

ID CATS_HUMAN STANDARD; PRT; 331 AA.
 AC P25774; Q9BUG3;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Cathepsin S precursor (EC 3.4.22.27).
 GN Name=CTSS;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae;
 OC Homo.
 OC NCBI_TaxId=9606;
 OK [1]
 RN NUCLEOTIDE SEQUENCE.
 RP TISSUE=Alveolar macrophage;
 RX MEDLINE=92218373; PubMed=1373132;
 RT Shi G.-P., Munger J.S., Meara J.P., Rich D.H., Chapman H.A.;
 RT "Molecular cloning and expression of human alveolar macrophage
 cathepsin S, an elastolytic cysteine protease.";
 RL J. Biol. Chem. 267:7258-7262(1992).
 [2]
 RN NUCLEOTIDE SEQUENCE.
 RP MEDLINE=94209337; PubMed=8157683;
 RX Shi G.-P., Webb A.C., Foster K.E., Knoll J.H.M., Lemere C.A.,
 RA Munger J.S., Chapman H.A.;
 RT "Human cathepsin S: chromosomal localization, gene structure, and
 tissue distribution.";
 RL J. Biol. Chem. 269:11530-11536(1994).
 [3]
 RN NUCLEOTIDE SEQUENCE [MRNA].
 RP TISSUE=Testis;
 RX MEDLINE=92317106; PubMed=1377692;
 RA Wiederanders B., Broemme D., Kirschke H., von Figura K., Schmidt B.,
 Peters C.;
 RT "Phylogenetic conservation of cysteine proteinases. Cloning and
 expression of a cDNA coding for human cathepsin S.";
 RL J. Biol. Chem. 267:13708-13713(1992).
 [4]
 RN SEQUENCE REVISION TO 211.
 RA Wiederanders B., Broemme D., Kirschke H., von Figura K., Schmidt B.,
 Peters C.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBD databases.
 [5]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RP TISSUE=Pancreas;
 RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strusberg R.L., Feingold B.A., Grouse I.H., Derge J.G.,
 Klausner R.D., Collins P.S., Wagner L., Shennan C.M., Schler G.D.,
 Altchul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stalcenon M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Pange C.,
 Raha S.S., Loggialano N.A., Peters G.J., Abramson R.D., Mulhany S.J.,
 Bosak S.A., McSwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultyk S.W.,
 Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marx M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [6]
 RN 3D-STRUCTURE MODELING OF 115-331.
 RX MEDLINE=99092748; PubMed=9876921; DOI=10.1093/protein/11.11.1007;
 RA Fengler A., Brandt W.;
 RT "Three-dimensional structures of the cysteine proteases cathepsins K
 and S deduced by knowledge-based modelling and active site
 characteristics.";

RL Protein Eng. 11:1007-1013(1998).
 CC -1- FUNCTION: Thiol protease. Key protease responsible for the removal
 CC of the invariant chain from MHC class II molecules. The bond-
 CC specificity of this proteinase is in part similar to the
 CC specificities of cathepsin L and cathepsin N.
 CC -1- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
 CC activity on Z-Phe-Arg-1-NHMe, and more activity on the Z-Val-Val-
 CC Arg-1-Xaa compound.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- SIMILARITY: Belongs to the peptidase C1 family.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 DR EMBL; S93414; AAB22005.1; -; mRNA.
 DR EMBL; M86553; AAB35655.1; -; mRNA.
 DR EMBL; U07374; AAB60643.2; -; Genomic DNA.
 DR EMBL; U07370; AAB60643.2; JOINED; Genomic DNA.
 DR EMBL; U07371; AAB60643.2; JOINED; Genomic DNA.
 DR EMBL; U07372; AAB60643.2; JOINED; Genomic DNA.
 DR EMBL; U07373; AAB60643.2; JOINED; Genomic DNA.
 DR EMBL; M90696; AAC37592.1; -; mRNA.
 DR EMBL; BC002642; AAH02642.1; -; mRNA.
 DR PIR; A42482; A42482.
 DR PDB; 1BXF; Model: A=115-331.
 DR PDB; 1GLO; X-ray; A=115-331.
 DR PDB; 1MS6; X-ray; A=115-331.
 DR PDB; 1NP2; X-ray; A/B=115-331.
 DR PDB; 1NQC; X-ray; A=115-331.
 DR MEROPS; C01.034; -;
 DR Ensembl; ENSG00000163131; Homo sapiens.
 DR HGNC; HGNC:32545; CTSS.
 DR H-InvDB; HIX0001035; -;
 DR MIM; 116845; -;
 DR GO; GO:0005576; C:extracellular region; NAS.
 DR GO; GO:0005764; C:lysosome; NAS.
 DR GO; GO:0004218; F:cathepsin S activity; TAS.
 DR GO; GO:0006955; P:immune response; TAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
 DR InterPro; IPR000169; Pept_cys_AS.
 DR InterPro; IPR000668; Peptidase_C1.
 DR PANTHER; PTHR12411; Peptidase_C1; 1.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN; Cl; 1.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE HIS; 1.
 KM 3D-structure; Glycoprotein; Hydrolase; Lysosome; Protease; Signal;
 FT Thiol protease; Zymogen.
 FT SIGNAL 1 16
 FT PROPEP 17 331
 FT CHAIN 115 331
 FT ACT_SITE 139 139
 FT ACT_SITE 278 278
 FT ACT_SITE 298 298
 FT CAROHYD 104 104
 FT DISULFID 126 224
 FT DISULFID 136 180
 FT DISULFID 170 213
 FT DISULFID 272 320
 FT DISULFID 92 92
 FT CONFLICT 113 113
 FT CONFLICT 161 161
 FT STRAND 119 120
 FT STRAND 121 124
 FT HELIX 125 125
 FT TURN 132 132
 FT STRAND 132 132
 M -> T (in Ref. 1 and 2).
 R -> W (in Ref. 5).
 T -> S (in Ref. 3 and 5).
 N-linked (GlcNAc...) (potential).
 Potential.
 Activation peptide.
 Cathepsin S.
 By similarity.
 By similarity.
 N-linked (GlcNAc...) (potential).

FT TURN 134 135
FT STRAND 137 137
FT HELIX 139 156
FT STRAND 162 162
FT HELIX 164 170
FT STRAND 173 175
FT TURN 176 176
FT STRAND 179 180
FT HELIX 181 181
FT STRAND 185 195
FT TURN 196 196
FT STRAND 198 200
FT TURN 201 203
FT HELIX 217 219
FT STRAND 220 222
FT TURN 226 229
FT TURN 232 233
FT HELIX 235 244
FT TURN 245 245
FT STRAND 248 252
FT HELIX 257 261
FT STRAND 266 271
FT TURN 270 271
FT STRAND 278 288
FT TURN 289 290
FT STRAND 291 297
FT TURN 300 300
FT TURN 302 303
FT TURN 305 305
FT STRAND 306 306
FT TURN 307 308
FT STRAND 309 313
FT HELIX 319 321
FT TURN 322 324
FT STRAND 327 331
SQ SEQUENCE 331 AA; 37510 MW; 835935FA56B78902 CRC64;

Query Match 99.5%; Score 1797; DB 1; Length 331;

Best Local Similarity 99.4%; Pred. No. 2.5e-14; Matches 329; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKRLVCLVLCSSAVALQHKDPTLDHMHLMKTYGKQYKKEKEBEAVRLIWEKNLKFVM 60
DB 1 MKRLVCLVLCSSAVALQHKDPTLDHMHLMKTYGKQYKKEKEBEAVRLIWEKNLKFVM 60
QY 61 LHNLEHSGMHSYDLGNHLDGMTSEEVMSLSLRVPSOMQRNITYKSNPNRILPDSVD 120
DB 61 LHNLEHSGMHSYDLGNHLDGMTSEEVMSLSLRVPSOMQRNITYKSNPNRILPDSVD 120
QY 121 WREKGCYTEVKYQSCGACMAFSAVGALEAOLKLTGKLVLSAONLVDCSTEKYGNKC 180
DB 121 WREKGCYTEVKYQSCGACMAFSAVGALEAOLKLTGKLVLSAONLVDCSTEKYGNKC 180
QY 181 NGGFMTTAFQYIIDNKIGIDSDASYPYKAMDLCQYDSKYRAATCSKYTELPYGRDVLKE 240
DB 181 NGGFMTTAFQYIIDNKIGIDSDASYPYKAMDLCQYDSKYRAATCSKYTELPYGRDVLKE 240
QY 241 AVANKGPVSVGVDAHPSPFLYRSGVYYPEPSCQNVNHNGLVVGVDLNGKEYMLVKNSW 300
DB 241 AVANKGPVSVGVDAHPSPFLYRSGVYYPEPSCQNVNHNGLVVGVDLNGKEYMLVKNSW 300
QY 301 GHNFGEGYITMARKNKNGHCGIASPSPYPEI 331
DB 301 GHNFGEGYITMARKNKNGHCGIASPSPYPEI 331

RESULT 3

ID Q6FHS5_HUMAN PRELIMINARY; PRT; 331 AA.
AC Q6FHS5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

DE CTSS protein.
GN Name-CTSS;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Halleck A., Ebert L., Moundinya M., Schick M., Eisenstein S.,
RA Neubert P., Ketrang K., Schatten R., Shen B., Henze S., Mar W.,
RA Korn B., Zuo D., Hu Y., Labaer J.,
RA Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -i SIMILARITY: Belongs to the peptidase C1 family.
DR EMBL: C941676; C946477.1; -, mRNA.
DR SMR; Q6FHS5; 115-331.
DR GO; GO:0004197; F:cyteine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000668; Peptidase C1.
DR InterPro; IPR000169; Pept. cys. AS.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPA1N.
DR ProDom; PD000156; Peptidase_C1; 1.
DR SMART; SM00645; Pept_C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW Hydrolyase; Protease; Thiol protease.
SQ SEQUENCE 331 AA; 37526 MW; 814FA33BDCL223B CRC64;

Query Match 99.2%; Score 1792; DB 2; Length 331;

Best Local Similarity 99.4%; Pred. No. 6.6e-14; Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCLVLCSSAVALQHKDPTLDHMHLMKTYGKQYKKEKEBEAVRLIWEKNLKFVM 60
DB 1 MKRLVCLVLCSSAVALQHKDPTLDHMHLMKTYGKQYKKEKEBEAVRLIWEKNLKFVM 60
QY 61 LHNLEHSGMHSYDLGNHLDGMTSEEVMSLSLRVPSOMQRNITYKSNPNRILPDSVD 120
DB 61 LHNLEHSGMHSYDLGNHLDGMTSEEVMSLSLRVPSOMQRNITYKSNPNRILPDSVD 120
QY 121 WREKGCYTEVKYQSCGACMAFSAVGALEAOLKLTGKLVLSAONLVDCSTEKYGNKC 180
DB 121 WREKGCYTEVKYQSCGACMAFSAVGALEAOLKLTGKLVLSAONLVDCSTEKYGNKC 180
QY 181 NGGFMTTAFQYIIDNKIGIDSDASYPYKAMDLCQYDSKYRAATCSKYTELPYGRDVLKE 240
DB 181 NGGFMTTAFQYIIDNKIGIDSDASYPYKAMDLCQYDSKYRAATCSKYTELPYGRDVLKE 240
QY 241 AVANKGPVSVGVDAHPSPFLYRSGVYYPEPSCQNVNHNGLVVGVDLNGKEYMLVKNSW 300
DB 241 AVANKGPVSVGVDAHPSPFLYRSGVYYPEPSCQNVNHNGLVVGVDLNGKEYMLVKNSW 300
QY 301 GHNFGEGYITMARKNKNGHCGIASPSPYPEI 331
DB 301 GHNFGEGYITMARKNKNGHCGIASPSPYPEI 331

RESULT 4

ID CATS_SAIBB STANDARD; PRT; 330 AA.
AC Q8HY82;
DT 10-MAY-2005 (Rel. 47, Created)
DT 10-MAY-2005 (Rel. 47, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Cachepoin S precursor (EC 3.4.22.27).
GN Name=CTSS;
OS Saimiri boliviensis boliviensis (Bolivian squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Platyrrhini; Cebidae;
OC Cebinae; Saimiri.
OX NCBI_TaxID=39432;

```

RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22538941; PubMed=12651112; DOI=10.1016/S1046-5928(02)00646-0;
RA Baker S.M., Karlsson L., Thurmond R.L.;
RT "Cloning, expression, purification, and activity of dog (Canis
  familiaris) and monkey (Saaimiri boliviensis) cathepsin S.";
RL Protein Expr. Purif. 28:93-101(2003).
CC -1- FUNCTION: Thiol protease. Key protease responsible for the removal
  of the invariant chain from MHC class II molecules. The bond-
  specificity of this proteinase is in part similar to the
  specificities of cathepsin L and cathepsin N (By similarity).
CC -1- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
  activity on Z-Phe-Arg-|-NHMeC, and more activity on the Z-Val-Val-
  Arg-|-Xaa compound.
CC -1- SUBCELLULAR LOCATION: Lysosomal (By similarity).
CC -1- SIMILARITY: Belongs to the peptidase C1 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use as long as its content is in no way modified and this statement is not
  removed.
CC -----
DR EMBL; AY156691; AA013008.1; -; mRNA.
DR HSSP; P25774; 1M56.
DR SMR; Q8HY82; 115-330.
DR MEROPS; C01.034; -.
DR InterPro; IPR000169; Pept_Cys_AS.
DR InterPro; IPR000668; Peptidase_C1.
DR PANTHER; PTHR12411; Peptidase_C1; 1.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPA1N.
DR PRODOM; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Pept_C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR GlycoProtein; HydroLase; Lysosome; Protease; Signal; Thiol protease;
  Zymogen.
KW SIGNAL.
FT PROPEP 1 17 Potential.
FT CHAIN 18 114 Activation peptide (By similarity).
FT ACT_SITE 115 330 Cathepsin S.
FT ACT_SITE 139 139 By similarity.
FT ACT_SITE 277 277 By similarity.
FT CARBOHYD 104 104 N-linked (GlcNAc...) (potential).
FT DISULFID 126 223 By similarity.
FT DISULFID 136 179 By similarity.
FT DISULFID 170 212 By similarity.
FT DISULFID 271 319 By similarity.
SQ SEQUENCE 330 AA; 37346 MW; 4ECE3129F2418C34 CRC64;

Query Match 94.2%; Score 1700.5; DB 1; Length 330;
Best Local Similarity 93.7%; Pred. No. 2.8e-133;
Matches 310; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

```

```

DB 240 AVANKGPVGVDSHSPFFLYRSGVYDPACTKRVNNGVLVIGDNLGKRYMLVKNSW 299
QY 301 GHNFGEGYIRMANRKNHCIAASFPSPYSEI 331
DB 300 GSNFGEGYIRMANRKNHCIAASFPSPYSEI 330

RESULT 5
CATS_CANFA
ID CATS_CANFA STANDARD; PRT; 331 AA.
AC O8HY81.
DT 10-MAY-2005 (Rel. 47, Created)
DT 10-MAY-2005 (Rel. 47, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Cathepsin S precursor (EC 3.4.22.27).
GN Name=CATS;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22538941; PubMed=12651112; DOI=10.1016/S1046-5928(02)00646-0;
RA Baker S.M., Karlsson L., Thurmond R.L.;
RT "Cloning, expression, purification, and activity of dog (Canis
  familiaris) and monkey (Saaimiri boliviensis) cathepsin S.";
RL Protein Expr. Purif. 28:93-101(2003).
CC -1- FUNCTION: Thiol protease. Key protease responsible for the removal
  of the invariant chain from MHC class II molecules. The bond-
  specificity of this proteinase is in part similar to the
  specificities of cathepsin L and cathepsin N (By similarity).
CC -1- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
  activity on Z-Phe-Arg-|-NHMeC, and more activity on the Z-Val-Val-
  Arg-|-Xaa compound.
CC -1- SUBCELLULAR LOCATION: Lysosomal (By similarity).
CC -1- SIMILARITY: Belongs to the peptidase C1 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use as long as its content is in no way modified and this statement is not
  removed.
CC -----
DR EMBL; AY156692; AA013009.1; -; mRNA.
DR HSSP; P25774; 1M56.
DR SMR; Q8HY81; 115-331.
DR MEROPS; C01.034; -.
DR Ensembl; ENSCAPG00000012086; Canis familiaris.
DR InterPro; IPR000169; Pept_Cys_AS.
DR InterPro; IPR000668; Peptidase_C1.
DR PANTHER; PTHR12411; Peptidase_C1; 1.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPA1N.
DR PRODOM; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Pept_C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR GlycoProtein; HydroLase; Lysosome; Protease; Signal; Thiol protease;
  Zymogen.
KW SIGNAL.
FT PROPEP 1 16 Potential.
FT CHAIN 17 114 Activation peptide (By similarity).
FT ACT_SITE 115 331 Cathepsin S.
FT ACT_SITE 139 139 By similarity.
FT ACT_SITE 278 278 By similarity.
FT ACT_SITE 298 298 By similarity.
FT CARBOHYD 104 104 N-linked (GlcNAc...) (potential).
FT DISULFID 126 224 By similarity.
FT DISULFID 136 180 By similarity.
FT DISULFID 170 213 By similarity.
FT DISULFID 272 320 By similarity.

```

Query Match	88.5%	Score 1599	DB 1	Length 331
Best Local Similarity	87.9%	Pred. No. 8-2e-125		
Matches 291	Conservative 19	Mismatches 21	Indels 0	Gaps 0
QY 1 MKRLVCVLLVCCSSAAVALHKKDPTLHHMHMLMKKTYGKQVKEKNBEAVRLIWEKMLKPYM 60				
DB 1 MKMLVGLPLDCCSYAAQVAKHDPTLHHMHMLMKKTYSKQVKEENBEAVRLIWEKMLKPYM 60				
QY 61 LHNLEHSMGMSHYDGLGMHLGDMTSEEVNLSMSLSLVPESQMRNITLYKSNPRLIPDSVD 120				
DB 61 LHNLEHSMGMSHYDGLGMHLGDMTSEEVNLSMSLSLVPESQMRNITLYKSNPRLIPDSVD 120				
QY 121 WREKCCATEVKYQSSCGACWAFSAVGALFAQLKLTGKLVLSLAQNLVDCSTERYKNKC 180				
DB 121 WREKCCATEVKYQSSCGACWAFSAVGALFAQLKLTGKLVLSLAQNLVDCSTERYKNKC 180				
QY 181 NGCFMTTAPQYIINDKIGDSDASYPKAMDLCQVDSKTRATCSKYTELPGREVDLKE 240				
DB 181 NGCFMTTAPQYIINDKIGDSDASYPKAMDLCQVDSKTRATCSKYTELPGREVDLKE 240				
QY 241 AVANKGPVSVGDADAHPSFFFLYRSQGYVEPSCQVNVNHHGLVVGADLNGKCYWLVKNSM 300				
DB 241 AVANKGPVSVADAHHSFFLYRSQGYVEPSCQVNVNHHGLVVGADLNGKCYWLVKNSM 300				
QY 301 GHNFGEGYIRMARANKGNHCGIASPSPYPEI 331				
DB 301 GHNFGEGYIRMARANKGNHCGIASPSPYPEI 331				
RESULT 6				
Q6PCUS_RAT .PRELIMINARY; PRT; 341 AA.				
AC 06PCUS5:				
DT 05-JUN-2004 (TREMBLERL. 27, Created)				
DT 05-JUN-2004 (TREMBLERL. 27, Last sequence update)				
DT 05-JUN-2004 (TREMBLERL. 27, Last annotation update)				
DE Ctes protein.				
OS Rattus norvegicus (Rat).				
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;				
OC Muridae; Murinae; Rattus.				
OX NCBI_TaxID=10116;				
RN [1]				
RP NUCLEOTIDE SEQUENCE.				
RC TISSUE=Pituitary gland;				
EX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;				
RA Straubeberg R.U., Felsingold E.A., Grouse L.H., Derge J.G.,				
RA Kleuser R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,				
RA Altschul S.F., Zeeberg B., Buecaw K.H., Scheffer C.F., Bhat N.K.,				
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,				
RA Diatchenko L., Matusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA Stepietson M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,				
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Plange C.,				
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,				
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,				
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA Rodriguez A.C., Truwood J., Schmutz J., Myers R.M.,				
RA Butterfield V.S., Krzyzinski M.I., Skalska U., Smalhus D.E.,				
RA Schnerch A., Schin J.E., Jones S.J.M., Marika M.A.,				
RT "Generation and initial analysis of more than 15,000 full-length human				
RT and mouse cDNA sequences.";				
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RP NUCLEOTIDE SEQUENCE.				
RC TISSUE=Pituitary gland;				
RA Straubeberg R.,				
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.				

Query	Match	77.3%	Score 1395.5	DB 2	Length 341
Best Local Similarity	77.5%	Pred. No. 7.9e-108			
Matches 255	Conservative	29	Mismatches 44	Indels 1	Gaps 1
CC	-1	Similarity: Belongs to the peptidase C1 family.			
DR	DR	BC059142; AAH59142.1; -, mRNA.			
DR	DR	HSPB; P43235; IAUU.			
DR	DR	SMR; O6PCU5; 124-341.			
DR	GO	GO:0004197; F:cyteline-type endopeptidase activity; IEA.			
DR	GO	GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro	IPR000668; Peptidase C1.			
DR	InterPro	IPR000169; Pept. cys. AS.			
DR	Pfam	PF00112; Peptidase_C1; 1.			
DR	PRINTS	PR00705; PAPA1N.			
DR	ProDom	PD000158; Peptidase C1; 1.			
DR	SMART	SM00645; Pept C1; 1.			
DR	PROSITE	PS00640; THIOI_PROTEASE ASN; 1.			
DR	PROSITE	PS00139; THIOI_PROTEASE_CYS; 1.			
DR	PROSITE	PS00639; THIOI_PROTEASE_HIS; 1.			
KW	Hydrolase	Protease; Thiol protease.			
SG	SEQUENCE	341 AA; 38573 MW; 833DFC8D58EC246 CRC64;			
QY	4	LVCVLVSSAAVLHKDPTLDHNNHMKTKYGGQYKKKNEAVRLIWEKNLKEVMLN 63			
DB	13	LFWPVPLVSVMEDQRPDLTDHMDLKKTHEKXYKQDNEDVARRLIWEKNLKEVMLN 72			
QY	64	LEHSMGMSYDLGNHLDGDMTSEVMSLMSLSLVPDSQMRNTYKSNPNRIIPDSVDR 123			
DB	73	LEHSMGMSYSVGNHNGMDMPEEVIGYMGSLIRIPRHNNRSGLTKSSNQTLPPDSVDR 132			
QY	124	KGCTTEVYKQSSCGACAFSAVAGLLEALDKLTGTLVLSLQNLVDCST-EKYGKKCG 182			
DB	133	KGCTVTVYKQSSCGSAFSAVAGLEGLKLTGTLVLSLQNLVDCSTEEYKNGKCG 192			
QY	183	GFMTTAFQYIIDNKGIDSDASYPYKAMDLCQYDSKYRAATCSKYTELPYGREVYLKEAV 242			
DB	193	GFMTTAFQYIIDNKGIDSDASYPYKAMDKECHYDKNRAATCSRIELPFGREVALKEAV 252			
QY	243	ANKGVSAGVDARHPSPFLYNSGVYEDPSTQNVMHGVLVYGGDLNGEKYVLVNSMGH 302			
DB	253	ATKGVSAGVDASHSFPFLYQSGVYDDPSTENVMHGVLVYGGDLNGEKYVLVNSMGH 312			
QY	303	NFGEBCYIRMARANKNHGCIASFSPSYPEI 331			
DB	313	HFGDQGYIRMARANKNHGCIASFSPSYPEI 341			
RESULT 7					
Q99M14 MOUSE					
ID	Q99M14	MOUSE PRELIMINARY;	PRT;	340 AA.	
AC	Q99M14;				
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)			
DE	Ctes Protein.				
GN	Name=Ctes;				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;				
OC	Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	11				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=CZECH 11;				
RC	TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;				
FX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;				
RA	Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,				
RA	Ditschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Schetz T.B.,				
RA	Brownstein M.J., Uadin T.B., Toshiyuki S., Carninci P., Prange C.,				

RA Raha S.S., Loquejano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., Mewen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Snailus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences",
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C2BECH II;
 RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
 RA Strusberg R.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1 - SIMILARITY: Belongs to the peptidase C1 family.
 DR EMBL; BC002125; AA02125.1; -; mRNA.
 DR HSSP; P25774; IM56.
 DR SMR; Q99M14; 123-340.
 DR MGI; MGI:107341; C1seb.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR GO; GO:0005764; C:lysosome; IDA.
 DR GO; GO:0016020; C:membrane; IDA.
 DR GO; GO:0004218; F:cathepsin S activity; IDA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IDA.
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; Pept_cys_AS.
 DR Pfam; PF00112; Peptidase_C1_1.
 DR PRINTS; PR00705; PAPA1N.
 DR ProDom; PD000158; Peptidase_C1_1.
 DR SMART; SM00645; Pept_C1_1.
 DR PROSITE; PS00640; TH1OL_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; TH1OL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; TH1OL_PROTEASE_HIS; 1.
 DR Hydrolase; Protease; Thiol protease.
 KW SEQUENCE 340 AA; 38456 MW; 07650261F319DB7 CRC64;
 SQ
 Query Match 75.5%; Score 1363.5; DB 2; Length 340;
 Best Local Similarity 75.6%; Pred. No. 3.7e-105;
 Matches 245; Conservative 37; Mismatches 41; Indels 1; Gaps 1;

AC Q8BSZ5;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Mus musculus 18 days pregnant adult female placenta and extra
 DE embryonic tissue cDNA, RIKEN full-length enriched library,
 DE clone:3830425101 product:cathepsin S, full insert sequence.
 GN Name=Ctss;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)00304-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
 RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055590;
 RA Kawai U., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Hara A., Fukunishi Y., Kono H., Aachi U., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuohi N., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quickenbush J.,
 RA Schiml L.M., Stäubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bash G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guernstein S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombauer P.,
 RA Nordone P., Ring B., Ringwald C., Rodriquez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsunk S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
 RA The FANTOM Consortium,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
 RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and substructure of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama U., Nishi K., Kitsuana T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara S., Wachiaki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai U.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format

RT sequencing pipeline with 384 multicapillary sequencer.;

RL Genome Res. 10:1757-1771 (2000).

RE NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;

RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,

RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,

RA Hayashida K., Hayata N., Hiramoto K., Hiraoka T., Hirozane T.,

RA Hori F., Imotani K., Iehi Y., Itoh M., Kagawa I., Kanakawa T.,

RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,

RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,

RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,

RA Saito R., Saitoh H., Sakai K., Sakai K., Sakazume N., Sano H.,

RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,

RA Tagawa A., Takahashi F., Takaku-Akhiba T., Takeda Y., Tanaka T.,

RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;

RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.

CC -1- SIMILARITY: Belongs to the peptidase C1 family.

DR EMBL; AK028366; BAC25906.1; -; mRNA.

DR HSSP; P25774; 1MS6.

DR SMR; O8B5Z5; 125-342.

DR MGI; MGI:107341; Ctes.

DR GO; GO:0005615; C:extracellular space; TAS.

DR GO; GO:0005764; C:Lysosome; IDA.

DR GO; GO:0016020; C:membrane; IDA.

DR GO; GO:0004218; F:cathepsin S activity; IDA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IDA.

DR InterPro; IPR000668; Peptidase_C1.

DR InterPro; IPR000169; Pept_cys_AS.

DR Pfam; PF00112; Peptidase_C1; 1.

DR PRINTS; PR00705; PAPAIN_.

DR PRODOM; PD000158; Peptidase_C1; 1.

DR SMART; SM00645; Pept_C1_1; 1.

DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.

DR PROSITE; PS00139; THIOL_PROTEASE CYS; 1.

DR PROSITE; PS00639; THIOL_PROTEASE HIS; 1.

KW Hydrolase; Protease; Thiol protease.

SC SEQUENCE 342 AA; 38707 MW; 0135655A819BA85 CRC64;

Query Match 75.5%; Score 1363.5; DB 2; Length 342;

Best local similarity 75.3%; Pred. No. 3.7e-105;

Matches 244; Conservative 38; Mismatches 41; Indels 1; Gaps 1;

QY 9 LVCSAAVAVQAKDPDLDHMHMKTKYKQYKKEKNEAVRRLIWKKNLKFVNLHLESM 68

DB 19 LVCSAVAMQQLQRPDLTDHMDWKTKHEKYDKNEEVRRLIWKKNLKFVNLHLEYSM 78

QY 69 GMSHYDLGNHLDGWTSEVNSIMSLRVPSCQMNITKYSNPNRILPDSVDMREKGCVT 128

DB 79 GMSHYDVGMDMGMQDNTNEILCRMGALRIIPRSPTVTIFRSVNTLTDYDMREKGCVT 138

QY 129 EVKTKGSCGACWAFSAVALERQALKTKGKLVLSAONLVDCST- EKYGNKCGNGEMTT 187

DB 139 EVKTKGSCGACWAFSAVALERQALKTKGKLVLSAONLVDCSNEEKYGNKCGCGYMT 198

QY 188 AFQYIINDKGIIDSDASPYRKAMDLCQYDSKTRATGSKYTELPRYGRDVLKEAANKGP 247

DB 199 AFQYIINDKGIIDSDASPYRKAMDLCQYDSKTRATGSKYTELPRYGRDVLKEAANKGP 258

QY 248 VSVGVDAHPSPFFLFRSGVYEPSCQVNVNHCGLVVGGLNGKGYMLVKNKSMGNFGE 307

DB 259 VSVGVDAHPSPFFLFRSGVYEPSCQVNVNHCGLVVGGLNGKGYMLVKNKSMGNFGE 318

QY 308 GYIMARKNKNHCGIASPSPYBEI 331

DB 319 GYIMARKNKNHCGIASPSPYBEI 342

RESULT 9

CATS_MOUSE STANDARD; PRT; 340 AA.

ID CATS_MOUSE

AC 070370; 054973;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DR 10-MAY-2005 (Rel. 47, Last annotation update)

DE Cathepsin S precursor (EC 3.4.22.27).

GN Name=Ctes; Synonyms=Cats;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RM [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=129/Sv, and BALB/c; TISSUE=Brain;

RA Doh-ura K.;

RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.

RM [2]

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Spleen;

RA Rommelskirch W.;

RM Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.

RM [3]

RP NUCLEOTIDE SEQUENCE OF 144-306, AND TISSUE SPECIFICITY.

RC STRAIN=C57BL/6; TISSUE=Cartilage;

RX MEDLINE=99326135; PubMed=10395917;

RA Soederstrom M., Salminen H., Glumoff V., Kirschke H., Aro H.,

RA Vuorio E.;

RT "Cathepsin expression during skeletal development.;"

RL Biochim. Biophys. Acta 1446:35-46(1999).

RM [4]

RP NUCLEOTIDE SEQUENCE OF 296-340.

RC STRAIN=C57BL/6; TISSUE=Brain;

RX MEDLINE=98184882; PubMed=9516475; DOI=10.1074/jbc.273.13.7691;

RA Dandoy-Dron F., Guillo F., Benoudjema L., Deslys J.-P., Laemess C.,

RA Dormont D., Tovey M.G., Dron M.;

RT "Gene expression in scrapie. Cloning of a new scrapie-responsive gene and the identification of increased levels of seven other mRNA transcripts.;"

RL J. Biol. Chem. 273:7691-7697(1998).

CC -1- FUNCTION: Thiol protease. Key protease responsible for the removal of the invariant chain from MHC class II molecules. The bond-specificity of this proteinase is in part similar to the specificities of cathepsin L and cathepsin N.

CC -1- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less activity on Z-Phe-Arg-|-NHMe, and more activity on the Z-Val-Val-Arg-|-Xaa compound.

CC -1- SUBCELLULAR LOCATION: Lysosomal.

CC -1- TISSUE SPECIFICITY: Widely expressed with highest expression found in non-skeletal tissues. Relatively high levels found in skeletal tissues.

CC -1- SIMILARITY: Belongs to the peptidase C1 family.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use as long as its content is in no way modified and this statement is not removed.

CC EMBL; AF051732; AAC05781.1; -; Genomic DNA.

CC EMBL; AF051727; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051728; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051729; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051726; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051730; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051731; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051732; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051733; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051734; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051735; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051736; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051737; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051738; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051739; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051740; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051741; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051742; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051743; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051744; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051745; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051746; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051747; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051748; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051749; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051750; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051751; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051752; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051753; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051754; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051755; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051756; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051757; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051758; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051759; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051760; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051761; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051762; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051763; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051764; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051765; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051766; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051767; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051768; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051769; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051770; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051771; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051772; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051773; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051774; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051775; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051776; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051777; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051778; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051779; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051780; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051781; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051782; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051783; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051784; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051785; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051786; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051787; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051788; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051789; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051790; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051791; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051792; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051793; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051794; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051795; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051796; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051797; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051798; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051799; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051800; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051801; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051802; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051803; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051804; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051805; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051806; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051807; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051808; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051809; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051810; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051811; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051812; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051813; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051814; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051815; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051816; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051817; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051818; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051819; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051820; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051821; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051822; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051823; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051824; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051825; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051826; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051827; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051828; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051829; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051830; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051831; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051832; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051833; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051834; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051835; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051836; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051837; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051838; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051839; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051840; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051841; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051842; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051843; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051844; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051845; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051846; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051847; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051848; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051849; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051850; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051851; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051852; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051853; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051854; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051855; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051856; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051857; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051858; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051859; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051860; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051861; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051862; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051863; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051864; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051865; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051866; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051867; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051868; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051869; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051870; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051871; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051872; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051873; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051874; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051875; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051876; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051877; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051878; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051879; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051880; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051881; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051882; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051883; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051884; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051885; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051886; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051887; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051888; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051889; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051890; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051891; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051892; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051893; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051894; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051895; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051896; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051897; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051898; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051899; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051900; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051901; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051902; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051903; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051904; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051905; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051906; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051907; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051908; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051909; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051910; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051911; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051912; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051913; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051914; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051915; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051916; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051917; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051918; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051919; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051920; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051921; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051922; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051923; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051924; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051925; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051926; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051927; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051928; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051929; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051930; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051931; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051932; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051933; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051934; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051935; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051936; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051937; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051938; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051939; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051940; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051941; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051942; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051943; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051944; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051945; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051946; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051947; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051948; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051949; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051950; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051951; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051952; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051953; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051954; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051955; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051956; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051957; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051958; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051959; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051960; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051961; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051962; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051963; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051964; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051965; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051966; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051967; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051968; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051969; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051970; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051971; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051972; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051973; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051974; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051975; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051976; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051977; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051978; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051979; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051980; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051981; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051982; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051983; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051984; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051985; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051986; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051987; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051988; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051989; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051990; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051991; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051992; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051993; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051994; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051995; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051996; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051997; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051998; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF051999; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF052000; AAC05781.1; JOINED; Genomic DNA.

CC EMBL; AF052001; AAC05781

DR GO; GO:0016020; C:membrane; IDA.
 DR GO; GO:0004218; F:cathepsin S activity; IDA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IDA.
 DR InterPro; IPR000169; Pept_Cys_AS.
 DR InterPro; IPR000668; Peptidase_C1.
 DR PANTHER; PTHR12411; Peptidase_C1; 1.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 DR 3D-structure; Glycoprotein; Hydrolase; Lysosome; Protease; Signal;
 KW Thiol protease; Zymogen.
 FT SIGNAL 1 17 Potential.
 FT PROPEP 18 122 Activation peptide (potential).
 FT CHAIN 123 340 Cathepsin S.
 FT ACT_SITE 147 147 By similarity.
 FT ACT_SITE 287 287 By similarity.
 FT ACT_SITE 307 307 By similarity.
 FT CARBOHYD 120 120 N-linked (GlcNAc...) (potential).
 FT DISULFID 134 233 By similarity.
 FT DISULFID 144 189 By similarity.
 FT DISULFID 178 222 By similarity.
 FT DISULFID 281 329 By similarity.
 FT VARIANT 218 218 T -> M.
 FT CONFLICT 1 29 MRAPGAARMLRMLPVCVAMRQLORD -> MAVLDAPG
 FT CONFLICT 34 34 Y -> H (in Ref. 2).
 FT CONFLICT 97 97 S -> L (in Ref. 2).
 FT CONFLICT 106 106 S -> P (in Ref. 2).
 FT CONFLICT 146 146 A -> S (in Ref. 3).
 SQ SEQUENCE 340 AA; 38438 MW; 068B61126E2E0C6 CRC64;

Query Match 74.9%; Score 1352.5; DB 1; Length 340;
 Best Local Similarity 75.0%; Pred. No. 3e-104;
 Matches 243; Conservative 37; Mismatches 43; Indels 1; Gaps 1;

QY 9 LVCSAVALQKHPDPLDHHMHMKTYGQYKKEKNEVRRLIMKKNLKFVNLHLSM 68
 DB 17 LVCSAVALQKHPDPLDHHMHMKTYGQYKKEKNEVRRLIMKKNLKFVNLHLSM 76
 QY 69 GMSYDLGMNLGDMTSEVSLMSLVPSCQWQNTIYKSNPNRILPDSYVWREKGCVT 128
 DB 77 GMSYDLGMNLGDMTSEVSLMSLVPSCQWQNTIYKSNPNRILPDSYVWREKGCVT 136
 QY 129 EYKYGSGCGACAFSAVAGLAEQQLKLGKLVLSAQNIVDCST-EKYGKCGCGGEMTT 187
 DB 137 EYKYGSGCGACAFSAVAGLAEQQLKLGKLVLSAQNIVDCST-EKYGKCGCGGEMTT 196
 QY 188 AFQYIIDKGDSDSYPRKAMDLCQYDSKYRATCSKYTELPYGRDVLKEAVANKGP 247
 DB 197 AFQYIIDKGDSDSYPRKAMDLCQYDSKYRATCSKYTELPYGRDVLKEAVANKGP 256
 QY 248 VSVGVDAHPSPFFLYRSYGVYEPSCQNVNMGVLVVGVDLNGKEVYLVKNSWGNFGE 307
 DB 257 VSVGVDAHPSPFFLYRSYGVYEPSCQNVNMGVLVVGVDLNGKEVYLVKNSWGNFGE 316
 QY 308 GYIRMAKRNKNGGGLASFPSTPEI 331
 DB 317 GYIRMAKRNKNGGGLASFPSTPEI 340

RESULT 10
 CATS_RAT STANDARD; PRT; 330 AA.
 AC Q02765;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Cathepsin S precursor (EC 3.4.22.27).
 GN Name=Ctbs;

OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Rattus.
 NC NCB1_TaxID=10116;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=93100327; PubMed=1281481;
 RA Petanceska S., Devil L.;
 *Sequence analysis, tissue distribution, and expression of rat
 cathepsin S.
 RL J. Biol. Chem. 267:26038-26043(1992).
 CC -1- FUNCTION: Thiol protease. Key protease responsible for the removal
 of the invariant chain from MHC class II molecules. The bond-
 specificity of this proteinase is in part similar to the
 specificities of cathepsin L and cathepsin N (By similarity). May
 be involved in thyroid hormone biosynthesis.
 CC -1- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
 activity on Z-Phe-Arg-|-NHMe, and more activity on the Z-Val-Val-
 Arg-|-Xaa compound.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- TISSUE SPECIFICITY: Highest levels occur in the ileum followed by
 spleen, brain, thyroid, ovary and uterus. Low levels are found in
 the liver, kidney, jejunum and lung with lowest levels in the
 heart.
 CC -1- INDUCTION: By thyroid-stimulating hormone.
 CC -1- SIMILARITY: Belongs to the peptidase C1 family.

CC EMBL; L03201; AAA40994.1; -; mRNA.
 CC PIR; A45087; A45087.
 CC HSSP; P25774; IM86.
 CC SMR; 002765; 113-330.
 CC MEROPS; C01.034; -;
 CC Ensembl; ENSRN000000021157; Rattus norvegicus.
 CC RGD; 621513; Ctbs.
 CC GO; GO:0004218; F:cathepsin S activity; IDA.
 CC InterPro; IPR000169; Pept_Cys_AS.
 CC InterPro; IPR000668; Peptidase_C1.
 CC PANTHER; PTHR12411; Peptidase_C1; 1.
 CC Pfam; PF00112; Peptidase_C1; 1.
 CC PRINTS; PR00705; PAPA1N.
 CC ProDom; PD000158; Peptidase_C1; 1.
 CC SMART; SM00645; Pept_C1; 1.
 CC PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 CC PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 CC PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 CC Glycoprotein; Hydrolase; Lysosome; Protease; Signal; Thiol protease;
 KW Zymogen.
 FT SIGNAL 1 17 Potential.
 FT PROPEP 18 122 Activation peptide.
 FT CHAIN 113 330 Cathepsin S.
 FT ACT_SITE 137 137 By similarity.
 FT ACT_SITE 277 277 By similarity.
 FT ACT_SITE 297 297 By similarity.
 FT CARBOHYD 100 100 N-linked (GlcNAc...) (potential).
 FT CARBOHYD 110 110 N-linked (GlcNAc...) (potential).
 FT DISULFID 124 222 By similarity.
 FT DISULFID 134 179 By similarity.
 FT DISULFID 168 211 By similarity.
 FT DISULFID 271 319 By similarity.
 SQ SEQUENCE 330 AA; 36833 MW; 670E3F08D7749BFE CRC64;

Query Match 72.0%; Score 1299.5; DB 1; Length 330;
 Best Local Similarity 76.9%; Pred. No. 7.6e-100;
 Matches 240; Conservative 28; Mismatches 41; Indels 3; Gaps 3;

QY 22 PTLDEHMHLMKKTGYKQYKENEAVRRLIWEKNTKFWLHNLBSHMGHSYDLGMNHLG 81
 DB 20 PTLDEHMHLMKKTGYKQYKENEAVRRLIWEKNTKFWLHNLBSHMGHSYDLGMNHLG 79
 QY 82 DMTSEVMSLSLRVPSQWQNRITKSNPNRILPDSVDMREKGCVTVEKTCQSGCAGWA 141
 DB 80 DMTPEVIGVMSLRIPRPMNRSGTLKSSNQTLDPDSVDMREKGCVTVEKTCQSGCAGWA 139
 QY 142 FSAVALLEAOLKLTNGKLVSLSAQNLNDGST-EKGNMGNGNGFMTTAPQYIINDKGLDS 200
 DB 140 FSAVALLEAOLKLTNGKLVSLSAQNLNDGST-EKGNMGNGNGFMTTAPQYIINDKGLDS 198
 QY 201 DASVYKAMDLCQYDSKYRATCSKYTELPYGRDEVLEKAVANKGPVSGV-DARHPSF 259
 DB 199 EASVYKAMDLCQYDSKYRATCSKYTELPYGRDEVLEKAVANKGPVSGV-DARHPSF 258
 QY 260 FLYRSGVYEPSCCTONVNHGVLVVGVDLNGEKYMLVKNMGNGEGBGYIMARANKGNH 319
 DB 259 FLYRSGVYEPSCCTONVNHGVLVVGVDLNGEKYMLVKNMGNGEGBGYIMARANKGNH 318
 QY 320 CGIASFSPSPYPI 331
 DB 319 CGIASFSPSPYPI 330

RESULT 11

Q5ZMR6_CHICK PRELIMINARY; PRT; 328 AA.
 AC Q5ZMR6;
 DT 25-OCT-2004 (TREMBLrel. 28, Created)
 DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=RCJMB04_1f23;
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus;
 CC NCBI_TaxID=9031;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=CB; TISSUE=Bursa;
 RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezubov Y., Zaim J.,
 RA Fiedler P., Kutter S., Blagoderaki A., Kostovska D., Kotter M.,
 RA Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.,
 RT "Full-length cDNAs from chicken bursa lymphocytes to facilitate
 RT gene function analysis."
 RL Genome Biol. 6:R6-R6(2005).
 CC -1- SIMILARITY: Belongs to the peptidase C1 family.
 DR EMBL, AJ719318; CAG30977.1; -, mRNA.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR InterPro; IPR000668; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000169; Pept. cys. AS.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE HIS; 1.
 KM Hydroxylase; Hypothetical protein; Protease; Thiol protease.
 SQ SEQUENCE 328 AA; 36397 MW; C5D007690421PE84 CRC64;

Query Match 67.4%; Score 1216.5; DB 2; Length 328;
 Best Local Similarity 67.4%; Pred. No. 6.3e-93;
 Matches 223; Conservative 36; Mismatches 69; Indels 3; Gaps 2;

QY 1 MKRLVCVLLVSSSAVALQHLKPTLDHMHLMKKTGYKQYKENEAVRRLIWEKNTKFWLHNL 60
 DB 1 MELLRCAVLLVTLVAVMGHPDPTLDHMHLMKKTGYKQYKENEAVRRLIWEKNTKFWLHNL 60

QY 61 LNLHLSHMGHSYDLGMNHLGDMTSEVMSLSLRVPSQWQNRITKSNPNRILPDSVD 120
 DB 61 LNLHLSHMGHSYDLGMNHLGDMTSEVMSLSLRVPSQWQNRITKSNPNRILPDSVD 118
 QY 121 WREKGCVTVEKTCQSGCAGWAPSAVGALEAOLKLTNGKLVSLSAQNLNDGST-EKGNMG 180
 DB 119 WREKGCVTVEKTCQSGCAGWAPSAVGALEAOLKLTNGKLVSLSAQNLNDGST-EKGNMG 177
 QY 181 NGFMTTAPQYIINDKGLDSASVYKAMDLCQYDSKYRATCSKYTELPYGRDEVLEK 240
 DB 178 NGFMTTAPQYIINDKGLDSASVYKAMDLCQYDSKYRATCSKYTELPYGRDEVLEK 237
 QY 241 AVANKGPVSGV-DARHPSF-FLYRSGVYEPSCCTONVNHGVLVVGVDLNGEKYMLVKN 300
 DB 238 AVANKGPVSGV-DARHPSF-FLYRSGVYEPSCCTONVNHGVLVVGVDLNGEKYMLVKN 297
 QY 301 GHNFGEBGYIMARANKGNHCGIASFSPSPYPI 331
 DB 298 GHNFGEBGYIMARANKGNHCGIASFSPSPYPI 328

RESULT 12

Q6DUC1_XENTR PRELIMINARY; PRT; 333 AA.
 AC Q6DUC1;
 DT 25-OCT-2004 (TREMBLrel. 28, Created)
 DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
 DE Cathepsin S.
 GN Name=ctss-prov;
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus; Silurana.
 CC NCBI_TaxID=8364;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Whole body;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stepieton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., USCIN T.B., Toshimiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Whole body;
 RA Klein S., Gerhardt D.S.,
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the peptidase C1 family.
 DR EMBL; BC075261; AAH75261.1; -, mRNA.
 DR Ensembl; ENSXETG0000008914; Xenopus tropicalis.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000169; Pept. cys. AS.
 DR InterPro; IPR000668; Peptidase_C1.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N.

DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept C1; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 KM Hydrolase; Protease; Thiol protease.
 SQ SEQUENCE 333 AA; 36839 MW; 65655F8A354407D CRC64;

Query Match 63.5%; Score 1147; DB 2; Length 333;
 Best Local Similarity 61.2%; Pred. No. 4e-87;
 Matches 205; Conservative 56; Mismatches 68; Indels 6; Gaps 3;

QY 1 MKRLVCLVLCSSAAVQAHKDPDLDHMLMKTKYTGKQYKKNBEAVRLLWEKQLKFM 60
 DB 1 MKSLIFLATVL-AAIVKARINPALNHLMLKNTNKKYEDIEDLQRIITWEKLANLVN 59
 QY 61 LHNLEHSGMHSYDGMNHLGDMTSEVWMSLSLRVPSQMRNTTYSNPRIT---LP 116
 DB 60 MENLEYSGMHTYELGMNHLADMTSEBKSKLTGLTPQSERQATFSQKSTFGKVP 119
 QY 117 DSVDRKQGVTEVYKQSCGACAFSAVGALEAQLKLTGVLVLSAQNIVDSTERYG 176
 DB 120 DSIDHRDGCUSVDVKNQGGCCSCAFSAVGALEGLMLKTGLVLSFQNLVDCCS-KYG 178
 QY 177 NKGCGFMFTTAFOYIIDNKGIDSDASYPYKAMDLCQYDSKRYATSKYTELPYGRD 236
 DB 179 NKGCGGFMFTAFOYIIDNKGIDSDASYPYHAMDEKCHDPTGKASTCAKYTEIIVPTED 238
 QY 237 VLKEVNAKGPVSVGDARHPFELRYSGVYRPSCTQNVNNGVLVNGYGLDNGEYLV 296
 DB 239 NLKQMLSGISGPISVAILDGTSPFELRYSGVYSDPTCSHEVNHGVLAVGYNLGDQFWLL 298
 QY 297 KNSMGNFGESEGYIMARNKGNHCGIASFPSPYPEI 331
 DB 299 KNSMGTXYGDOGYVARIANKGLCGVASTCTCPPEI 333

RESULT 13
 Q6DE57 XENLA PRELIMINARY; PRT; 320 AA.

ID 06DE57 XENLA PRELIMINARY; PRT; 320 AA.
 AC 06DE57
 DT 25-OCT-2004 (TREMBlrel. 28, Created)
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
 DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
 DE Ctes-prov protein.
 GN Name=ctes-prov;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus; Xenopus.
 OC NCBI_TaxId=8355;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Splicein;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Klausner R.L., Pelngold E.A., Grouse L.H., Derge J.G.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stajichon M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaeetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.U.,
 RA Boses S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.N.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalski U., Smallos D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RM [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Splicein;
 RA Klein S., Gerhard D.S.,
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 CC -1. SIMILARITY: Belongs to the peptidase C1 family.
 DR EMBL; BC077285; AAH77285.1; -; mRNA.
 DR GO; GO:0004197; F:cytosine-type endopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; Pept_cys_AS.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept C1; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 KM Hydrolase; Protease; Thiol protease.
 SQ SEQUENCE 320 AA; 35654 MW; 1D7C3DC9AD433FA CRC64;

Query Match 62.3%; Score 1124.5; DB 2; Length 320;
 Best Local Similarity 63.2%; Pred. No. 2.9e-85;
 Matches 199; Conservative 50; Mismatches 61; Indels 5; Gaps 2;

QY 21 DPTLDHMLMKTKYQKQYKKNBEAVRLLWEKQLKFMVNLNLEHSGMHSYDGMNHL 80
 DB 7 DPALDNLMLKMKTKYKTEDEDEDLRLITWEKLANLVNNGMHTYELGMNHL 66
 QY 81 GDMTSEVWMSLSLRVPSQMRNTTYSNPRIT---PSVDRKQGVTEVYKQSC 136
 DB 67 ADMTSEBKSKMTGLTPHSEKATFSQKSTLGGKVPSPSIDREKGVSEVKNQGC 126
 QY 137 GACAFSAVGALEAQLKLTGKLVLVLSAQNIVDSTERYGKNGCGFMFTTAFOYIIDN 196
 DB 127 GSCAFSAVGALEGLMLKTGLVLSFQNLVDCCS-KYKNGCGFMFTTAFOYIIDN 185
 QY 197 GIDSASYPYKAMDLCQYDSKRYATSKYTELPYGRDVLKEAVANKGPVSGVDARH 256
 DB 186 GIDSPTYPYHAMDEKCHYELAGKASCVKRYEIVPTEDMLKQMLGNIGPISVALDGT 245
 QY 257 PSFELRYSGVYRPSCTQNVNNGVLVNGYGLDNGEYLVNNSMGNFGESEGYIMARNK 316
 DB 246 PTFELRYSGVYSDPTCSHEVNHGVLAVGYNLGDQFWLLKNSMGTXYGDOGYVARIANK 305
 QY 317 GNHCGIASFPSPYPEI 331
 DB 306 ENLCGVASTYSPPEI 320

RESULT 14
 Q7T0S4 XENLA PRELIMINARY; PRT; 333 AA.
 ID 07T0S4 XENLA PRELIMINARY; PRT; 333 AA.
 AC 07T0S4
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Ctes-prov protein.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus; Xenopus.

OK NCBI_TaxID=8355;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Spleen;
 RX MEDLINE=2238857; PubMed=1247932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zolbins B., Buetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh L.,
 RA Dichtchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ussdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Boeck S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Hellon E., Kettlemen M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalski U., Smallos D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [2]
 RN NUCLEOTIDE SEQUENCE.
 RC TISSUE=Spleen;
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative.";
 RL Dev. Dyn. 225:384-391 (2002).
 [3]
 RN NUCLEOTIDE SEQUENCE.
 RC TISSUE=Spleen;
 RA Klein S., Strausberg R.;
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the peptidase C1 family.
 DR EMBL; BC056059; AAB56059.1; -; mRNA.
 DR HSSP; P25774; INOC.
 DR MEROPS; C01.034; -;
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000668; Peptidase_C1.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N_1.
 DR PRODOM; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 KW Hydrolase; Protease; Thiol protease.
 SQ SEQUENCE 333 AA; 36943 MW; 655534D5084059FA CRC64;
 Query Match 61.4%; Score 1109.5; DB 2; Length 333;
 Best Local Similarity 58.9%; Pred. No. 5.4e-84;
 Matches 196; Conservative 61; Mismatches 71; Indels 5; Gaps 2;
 QY 3 RLVCVLAVSSAVVAALHDKPTLDHMHMLKTKYGYKQYKKEAEAVRLIWEKQLKFWMLN 62
 DB 2 RSLFLLALATLTYVARINPDLNMLKNTKSHKEYTEETDLDRLIWEKQLDLPVNNH 61
 QY 63 NLEHSMGNHSDYLDGNHLDGMTSEEVMSLSLVRPSQWQRNITKYSNPRI---LPDS 118
 DB 62 NLEYSMGHTYELGNHLDGMTSEEMKSLTGLILPPISEKAKRSQGRNPGSKVADS 121
 QY 119 VDMREKGVTEYKYVYGGSCGACAFSAVGALEKQLKTKGKVLVSAQNLVDCSTKRYGNK 178
 DB 122 IDMRDKGVSDYKNGGCGSCWAFSAVGALEKQLKTKGKVLVSAQNLVDCSTKRYGNK 180

QY 179 GCNGGFMTAFQYIINDNKIGSDASYPYKAMDKCQYDSTKRAATCSKYTELPYREDVL 238
 DB 181 GCGGFMTSAFOYIYDNNNGIDSDSYYPYHAMDKECHYELAGRASSCVKXTEIVPCEINVL 240
 QY 229 KEAVANKPVSVDARPSFPLRSQYIYEPSCQNVNNGVLYVVGDLNGKEFWLVKN 298
 DB 241 KQALGTGIPISVAIDGTPTPEFLKYSQVSPSCSQEVNHGVLAIYGTLNGQDFMLKN 300
 QY 299 SMGNHFGEGYIRMAARKNGHGTIASPSPYBEI 331
 DB 301 SWGTYIGDKGFVRIRKNGNLGVASTYSTPEI 333
 RESULT 15
 ID Q7T183_9C1CH PRELIMINARY; PRT; 334 AA.
 AC Q7T183;
 DT 01-OCT-2003 (TREMBLrel. 25, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Cathepsin.
 OS Parabolidochromis chilotae.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;
 OC Cichlidae; African cichlids; Pseudocrenilabrinae; Haplochromini;
 OC Parabolidochromis.
 OK NCBI_TaxID=77306;
 [1]
 RN NUCLEOTIDE SEQUENCE.
 RC TISSUE=Head;
 RX MEDLINE=22871689; PubMed=14507309;
 RL DOI=10.1046/j.1365-3083.2003.01332.x;
 RA Uinuk-ool T.S., Takezaki N., Kuroda N., Figueroa F., Sato A.,
 RA Samonte I.E., Mayer W.E., Klein J.;
 RT "Phylogeny of antigen-processing enzymes: cathepsins of a
 RT cephalochordate, an agnathan and a bony fish.";
 RL Scand. J. Immunol. 58:436-448 (2003).
 [2]
 RN NUCLEOTIDE SEQUENCE.
 RC TISSUE=Head;
 RA Uinuk-ool T.S., Takezaki N., Kuroda N., Figueroa F., Sato A.,
 RA Samonte I.E., Mayer W.E., Klein J.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the peptidase C1 family.
 DR EMBL; AY333300; AAQ01147.1; -; mRNA.
 DR HSSP; P25774; INOC.
 DR MEROPS; C01.034; -;
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000668; Peptidase_C1.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N_1.
 DR PRODOM; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
 KW Hydrolase; Protease; Thiol protease.
 SQ SEQUENCE 334 AA; 36543 MW; A227955CB423E537 CRC64;
 Query Match 60.1%; Score 1086; DB 2; Length 334;
 Best Local Similarity 60.9%; Pred. No. 4.9e-82;
 Matches 199; Conservative 40; Mismatches 84; Indels 4; Gaps 3;
 QY 4 LVCTLVSSAVVAALHDKPTLDHMHMLKTKYGYKQYKKEAEAVRLIWEKQLKFWMLN 63
 DB 9 LLILLISLCAVAAAF--ESTIDAHMELMKTKHGSYKNDVNAHRELMGNLKNLITVHN 66
 QY 64 LEHSMGNHSDYLDGNHLDGMTSEEVMSLSLVRPSQWQRNIT-KYSNPRIPLDPSVDMR 122
 DB 67 LEASGMLHTYELGNHLDGMTSEELIMQFFASLTPTPTDIDRAPSPPAGASGSIPTDMWR 126

RESULT 2
US-10-995-561-978
; Sequence 978, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C1001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 978
; LENGTH: 331
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-995-561-978

Query Match 99.7%; Score 1800; DB 6; Length 331;
Best Local Similarity 99.7%; Pred. No. 3.2e-160;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MKRLVCVLLVCSSAAVAQJHKDPTLDHMHLMKTKYQYKEKNEAVRLLIWEKXIKFVM 60
DB 1 MKRLVCVLLVCSSAAVAQJHKDPTLDHMHLMKTKYQYKEKNEAVRLLIWEKXIKFVM 60
QY 61 LHNLEHSMGMSYDIGNHLDGMTSEEVMSLSLRVPSQORNTTYSNPRILIPDSVD 120
DB 61 LHNLEHSMGMSYDIGNHLDGMTSEEVMSLSLRVPSQORNTTYSNPRILIPDSVD 120
QY 121 WREKCVTEVKKYQSGCGMAFSAVGALBAQJLKTGKLVSLAQNLYDCSTKXGNKGC 180
DB 121 WREKCVTEVKKYQSGCGMAFSAVGALBAQJLKTGKLVSLAQNLYDCSTKXGNKGC 180
QY 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKTYRAATCSKYTELPGREDVJKE 240
DB 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKTYRAATCSKYTELPGREDVJKE 240
QY 241 AVANTGPVSVGVADARHPSFFLYRSGYTYRPSCTONVNGVLYVGGDLNGKRYLVKXSW 300
DB 241 AVANTGPVSVGVADARHPSFFLYRSGYTYRPSCTONVNGVLYVGGDLNGKRYLVKXSW 300
QY 301 GHNFGSEGYIRMARNGKNGHCGIASPSPYPEI 331
DB 301 GHNFGSEGYIRMARNGKNGHCGIASPSPYPEI 331

RESULT 3
US-10-995-561-976
; Sequence 976, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C1001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 976
; LENGTH: 220
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-995-561-976

Query Match 62.5%; Score 1129; DB 6; Length 220;
Best Local Similarity 100.0%; Pred. No. 4.8e-98;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVCSSAAVAQJHKDPTLDHMHLMKTKYQYKEKNEAVRLLIWEKXIKFVM 60
DB 1 MKRLVCVLLVCSSAAVAQJHKDPTLDHMHLMKTKYQYKEKNEAVRLLIWEKXIKFVM 60
QY 61 LHNLEHSMGMSYDIGNHLDGMTSEEVMSLSLRVPSQORNTTYSNPRILIPDSVD 120
DB 61 LHNLEHSMGMSYDIGNHLDGMTSEEVMSLSLRVPSQORNTTYSNPRILIPDSVD 120
QY 121 WREKCVTEVKKYQSGCGMAFSAVGALBAQJLKTGKLVSLAQNLYDCSTKXGNKGC 180
DB 121 WREKCVTEVKKYQSGCGMAFSAVGALBAQJLKTGKLVSLAQNLYDCSTKXGNKGC 180
QY 181 NGGFMTTAFQYIINDKIGDSASYPYKAM 209
DB 181 NGGFMTTAFQYIINDKIGDSASYPYKAM 209

RESULT 4
US-10-131-826A-12
; Sequence 12, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; PRIOR FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 12
; LENGTH: 334
; TYPE: PR1
; ORGANISM: Homo Sapien
US-10-131-826A-12

Query Match 47.0%; Score 849; DB 6; Length 334;
 Best Local Similarity 49.8%; Pred. No. 9.2e-72;
 Matches 166; Conservative 55; Mismatches 102; Indels 10; Gaps 6;

4 LVCLVLVSSSAVAQHLHPTLDHMHMLKTKYKQYKKEAEVARRLWENKLEFVWLHN 63
 7 LAACFLGJASAVPKF--DQNLDTKYOQWKATHRLLY--GANEGBGMRRAWEKMKMIELHN 63

64 LEHSMGMSYDLGNMHLGDMTSEEVWSLMSLRVPSQMRNITYKSNPNRLLPDSVDWRE 123
 64 GEYQGGKGFMTMANNAFEDMTNEEROMGCFR--NQRKRKGVFRPEPLFLDLPSSVDMRK 122

124 KGVTEVYKQSCGACMAFSAVGALEAQLKLTGKGLVSLAQNVLVDCSTERYKNGKNGG 183
 123 KGYVTPVKNQKCCSCMAFSAATGALGEMFRKTGLVLSRQNLVDCSRPQ--GNQCGNGG 181

184 FMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKRYAATSKYTELPYGEEDVLKXVA 243
 182 FMARAFQYVKNGGIGDSSESYPPYAVDEICRYRPENSVANDTGFVVAAPGEKALMKAVA 241

244 NKGPVSGVDARHPSFFLYRSQVYEPSC--TONVNHGVLVVG---DLNGKEYWLKYN 298
 242 TVGPISTVMDHSHSFQYKSGITFEPDCCSRNLDHGLVVGIGFEGANSNNKTYLVKN 301

299 SMGHNFGEGYIRMANRKNHCGIASFPSPYPI 331
 302 SMGPWMSGNGYKIAKDNKNGIATATASYPNV 334

RESULT 5
 US-10-821-234-1672
 ; Sequence 1672, Application US/10821234
 ; Publication No. US20050255114A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Labat, Ivan
 ; APPLICANT: Stache-Crain, Birgit
 ; APPLICANT: Andarmant, Susan
 ; APPLICANT: Tang, Y. Tom
 ; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
 ; FILE REFERENCE: 821A
 ; CURRENT APPLICATION NUMBER: US/10/821,234
 ; CURRENT FILING DATE: 2004-04-07
 ; PRIOR APPLICATION NUMBER: US 60/462,047
 ; PRIOR FILING DATE: 2003-04-07
 ; NUMBER OF SEQ ID NOS: 1704
 ; SOFTWARE: pc_seq_genes Version 1.0
 ; SEQ ID NO 1672
 ; LENGTH: 333
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-821-234-1672

Query Match 46.5%; Score 840.5; DB 6; Length 333;
 Best Local Similarity 48.5%; Pred. No. 5.7e-71;
 Matches 164; Conservative 57; Mismatches 98; Indels 19; Gaps 8;

4 LVCLVLVSSSAVAQHLHPTLDHMHMLKTKYKQYKKEAEVARRLWENKLEFVWLHN 63
 5 LILAFCLGJASATLTFPHSLDAQWTWKAKMHNRLY--GANEGBGMRRAWEKMKMIELHN 63

64 LEHSMGMSYDLGNMHLGDMTSEEVWSLMSLR--RVPSQ---WQRNITYKSNPNRLLPDS 118
 64 GEYQGGKGFMTMANNAFEDMTSEEROMNGCFR--NQRKRKGVFRPEPLFLDLPSSVDMRK 117

119 VDMREKGVTEVYKQSCGACMAFSAVGALEAQLKLTGKGLVSLAQNVLVDCSTERYKNGK 178
 118 VDMREKGVTEVYKQSCGACMAFSAATGALGEMFRKTGLVLSRQNLVDCSRPQ--GNE 176

179 GONGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKRYAATSKYTELPYGEEDVLKXVA 238
 177 GONGGLMAYAFQYVODNGLDSEESYPYEAATESCKYKPKSVANDTGFVDP--KQEKAL 235

239 KEAVANKGPVSGVDARHPSFFLYRSQVYEPSC--TONVNHGVLVVG---DLNGKEY 293

236 MKAAVATGPISTVADAGHESFLYKGEYFEPDCCSDEMDHGVLVVGGFESTSDNNKY 295
 294 WLKNSWGNHNGEGYIRMANRKNHCGIASFPSPYPI 331
 296 WLKNSWGEWGMGIVYAKADRNHCGIASAASYPV 333

RESULT 6
 US-10-995-561-866
 ; Sequence 866, Application US/10995561
 ; Publication No. US20050272054A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cargill, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF
 ; FILE REFERENCE: C0001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561
 ; CURRENT FILING DATE: 2004-11-24
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: FastSeq for windows Version 4.0
 ; SEQ ID NO 866
 ; LENGTH: 335
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-995-561-866

Query Match 29.7%; Score 536.5; DB 6; Length 335;
 Best Local Similarity 37.9%; Pred. No. 1.1e-42;
 Matches 125; Conservative 49; Mismatches 129; Indels 27; Gaps 12;

10 VCSA---VAQHLHPTLDHMHMLKTKYKQYKKEAEVARRLWENKLEFVWLHN 65
 19 VCGAAELCVNSLEK----HFKSMGSKRRTY--STREYHRLQTPASMRKINAH-- 69

66 HSMGMSYDLGNMHLGDMTSEEVWS--LMSLRVPSQMRNITYKSNPNRLLPDSVDWRE 123
 70 --NGNHFPRKALNPFSDMSFAIRIGHKLYMSEPRQNSATKSNYLVGTGP--YPSVDWPK 124

124 KGVTEVYKQSCGACMAFSAVGALEAQLKLTGKGLVSLAQNVLVDCSTERYKNGKNGG 182
 125 KGNVSPVKNQKCCSCMAFSAATGALGEMFRKTGLVLSRQNLVDCSRPQ--GNQCGNGG 183

183 GFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKRYAATSKYTELPYGEEDVLKXVA 242
 184 GLPSQAFETIILYNGIMGEDTYPYQKQYCKFPQKALGFVKOVANITTYDEEAMVEAV 243

243 ANKGPVSGVDARHPSFFLYRSQVYEPSC--TONVNHGVLVVGGLDNGKEYWLKYN 299
 244 ALVNPVSAFPEVQ--DFWMTYRTGYSSTSCIKTDPKNHVLAVGIGBKNGIPYWIYVNS 302

300 WGNHFGEGYIRMANRKNHCGIASFPSPY 329
 303 WGPQMGNGYFLIRGK--NMGCLAACASYP 331

RESULT 7
 US-10-821-234-1211
 ; Sequence 1211, Application US/10821234
 ; Publication No. US20050255114A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Labat, Ivan
 ; APPLICANT: Stache-Crain, Birgit
 ; APPLICANT: Andarmant, Susan
 ; APPLICANT: Tang, Y. Tom
 ; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
 ; FILE REFERENCE: 821A
 ; CURRENT APPLICATION NUMBER: US/10/821,234
 ; CURRENT FILING DATE: 2004-04-07
 ; PRIOR APPLICATION NUMBER: US 60/462,047
 ; PRIOR FILING DATE: 2003-04-07
 ; NUMBER OF SEQ ID NOS: 1704

Db 273 PVRNAGSGSCSYFASMGMLERIRILITNNSOTPLISQEVVSGS--QYA-QGCGSG--- 326
QY 187 TAFQYIINK-----GISDASYPYKAMDLCQ-----YDKYRATSCKYTELPGY 233
Db 327 --FPLINGKTAODPGLVEACFPYTGDSPCMKEDCFRYSSEYH-----YVGGFYG 378
QY 224 --REVLEAVANKGPVGVDAHPSFPLRSGVYEPSCF-----QNVNHYGLVNGY 285
Db 379 GCNEALMLTELVEHGMVAFAFEV-YDDEFLHYKKGITHTGLRDPNPELTNHAVALVGY 437
QY 286 G--DLNGEKYWLVKNSMGHNFGSEGYIRMARKNGHCGIAS 324
Db 438 GTDSASGMDWMLVKNSMGHNGENGYFPIRIGCT-DECAIES 477

RESULT 10
US-11-102-883-20
; Sequence 20, Application US/11102883
; Publication No. US20050281816A1

; GENERAL INFORMATION:
; APPLICANT: Lamping, Norbert
; APPLICANT: Cramer, Reto
; APPLICANT: Fluckiger, Sabina
; APPLICANT: Daigle, Isabelle
; TITLE OF INVENTION: Modular Antigen Transporter Molecules (MAT Molecules) for
; TITLE OF INVENTION: Modulating Immune Reactions, Associated Constructs, Methods and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 03100234pa
; CURRENT APPLICATION NUMBER: US/11/102,883
; PRIOR FILING DATE: 2005-04-11
; PRIOR APPLICATION NUMBER: EP02022774.0
; PRIOR FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: PCT/EP2003/011190
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patent version 3.2
; SEQ ID NO 20
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Dermotophagoides pteronyssinus
US-11-102-883-20

Query Match 17.6%; Score 318.5; DB 7; Length 320;
Best Local Similarity 28.0%; Pred. No. 1.9e-22;
Matches 91; Conservative 66; Mismatches 129; Indels 39; Gaps 12;

QY 7 VLVCS-SAVAQIHKDPTLDHMHMLKTYGKQYK-KNEBAVRRLIWEKNLKFVMLNL 64
Db 4 VLAISLALASVAVARPSIKTFEEYKKAFNKSVATFDEBARKNPFE-SVKYV----- 57
QY 65 EHSWGMHSYDGMNHLGDMTSEE-----VMSIMSLRVPSCQORNI-TYKSNPNRILPDS 118
Db 58 -----QNGGCAINHLSDLSLDEFKXRFILMSAEAFEHKTPDLDLAETNACISINGNAPAE 111
QY 119 VDMREKGVTVKXGSGCAWAFSAVAGLEQLTKTGKIVLSAQNVLVDSTEKYGNK 178
Db 112 IDLRMRTVTPIRMQGGSCGSCAFSGVATESAYLATRNQSIDLEQELVDCASO-----H 167
QY 179 GCNGGFMTTAFQYIINKGIDSDASYPYKAMDLCQYDSKYRAATSCKYTELPGYREBYL 238
Db 168 GCHGHTIRGIEYIGHN-GVVEGSEYRYVARQSCRPAHQFQ-FG-INSYCOIYPNNWKI 225
QY 229 KEAVA-NGEPVSVGDARHPSEFFLYRSGY-----YEPSTQVNVHGVLVGGLDNG 290
Db 226 REALQHTSAIAVIGIIDLDAFRHYDGRITIIORDNGYQPN-----YHAVNIVGYSNQG 280
QY 291 KEYWLKNSMGHNFGSEGYIRMARKNGHCGIAS 315
Db 281 VDYWLKNSMGHNFGSEGYIRMARKNGHCGIAS 305

RESULT 11

US-10-878-556A-18
; Sequence 18, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
; APPLICANT: Hoffmann La-Roche Inc.
; TITLE OF INVENTION: HCV regulated protein expression
; FILE REFERENCE: 21762
; CURRENT APPLICATION NUMBER: US/10/878,556A
; CURRENT FILING DATE: 2004-06-28
; NUMBER OF SEQ ID NOS: 199
; SOFTWARE: Patent version 3.1
; SEQ ID NO 18
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ENTRY DATE: 1988-08-01
US-10-878-556A-18

Query Match 17.1%; Score 309; DB 6; Length 339;
Best Local Similarity 25.3%; Pred. No. 1.6e-21;
Matches 92; Conservative 57; Mismatches 123; Indels 92; Gaps 13;

QY 5 VCVLLVCSAAVAQIHKDPTLDHMHMLKTYGKQYK-KNEBAVRRLIWEKNLKFVMLNL 64
Db 8 LCCLLVLANARSRSFHFVSD-----ELVNVYKQNT 39
QY 65 EHSWGMHSYDGMNHL-----GDMTSEEVMSLMSLRVPSCQORNIITYKSNPNRIL 115
Db 40 TWOAGHNFYNDVMSLYKLCTPLGGRPPORVM-FTEDLTKLPASFDR----- 87
QY 116 PDSVDMREKGVTVKXGSGCAWAFSAVAGLEQLTKTGKIVLSAQNVLVDSTEKYGNK 173
Db 88 --EQMPCPTTKETRDGSGSCWAFSAVAGLEQLTKTGKIVLSAQNVLVDSTEKYGNK 144
QY 174 KYGNKGCNGGFMTTAFQYIINKGIDSDASY-----PYKAMDLCQYDSKYRAAT--- 223
Db 145 MCGD-GCNGGYPAEAKMN-WTRKGLVSGGLYESHVGCPRYSIPPEHHVNGSRPCTGEG 202
QY 224 ---CSKYTELPGY-----GREVLEAVANKGPVSVGDARHPSEFF 260
Db 203 DTPKSKICEPYSPTYKQDKHYGNSYSVSNSEKIDIAIYKN-GPYE-GAFSYSDPL 260
QY 261 LYRSGVYEPSCQVNVHGVLVGGLDNGEKYWLKNSMGHNFGSEGYIRMARKNGHCG 320
Db 261 LYKSGVYOHVTEGEMWGHAIIRILGWEVNGTTPYMLVANSWNTDWDGNGFFKILRQ-DHC 319
QY 321 GIAS 324
Db 320 GIES 323

RESULT 12
US-10-995-561-681
; Sequence 681, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: Cargili, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 681
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-681

;; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;; TITLE OF INVENTION: DETECTION AND USES THEREOF
;; FILE REFERENCE: C1001559
;; CURRENT APPLICATION NUMBER: US/10/995,561
;; CURRENT FILING DATE: 2004-11-24
;; NUMBER OF SEQ ID NOS: 85702
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 685
;; LENGTH: 339
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-995-561-685

Query Match 17.1%; Score 308; DB 6; Length 339;
Best Local Similarity 25.3%; Pred. No. 28-21; Indels 92; Gaps 13;

Matches 92; Conservative 57; Mismatches 123; Indels 92; Gaps 13;

QY 5 VCVLLVCSAQAQLHDKDPTLDHMHMLWKTYGQYKEKNEAVRRLIWEKNLKFVMLHNL 64
DB 8 LCCLLVLANARSRPSFHPISD-----ELVNVYVKRNT 39
QY 65 EHSWGMHSYDLAGMHL-----GDMTSEVMSLMSLRVPSQOMQRNITYKSNPNRIL 115
DB 40 TWQAGHNFPYNVDMSYLKRLCGTFGLGPKPPQPRVM-FTBDLKLPAASFDR----- 87
QY 116 PDSVDMREKGCVTEYKYGSSGCAWAFSAVGALEAQLKLTGKLYS--LSAQNLYDCSTE 173
DB 88 --EQWPCPTLKEIRDOSSCGSCWAFGAVEAISDRICHTNAHVSVEVSADLLTCCGS 144
QY 174 KYGNKGCNGGFWMTAFQYIIDNKGIDSDASY-----PYKAMDLCQYDSKYRAAT--- 223
DB 145 MCGD-GCNGGYPAEAMNF-WTRKGLVSGGLYSHVGCRRYSIPCEHHVNGSRPCTGEG 202
QY 224 ---CSKYTELPY-----GREVLKEAVANKGPVSVGVDAHPSPF 260
DB 203 DTPKCSKICEPQSPPTYKODKHGYNSYSVSNSEKDIIMAEIYKN-GPVE-GAFSVYSDFL 260
QY 261 LYRSGVYVEPSCTONVNHGVLVVGVDLNGKKEWMLVKNSWGHNPGEBGYIRMARKNHC 320
DB 261 LYKSGVYGHVTEGMMGHAIRILGWGVENGTPYWLAVNSMNTDWDGNGPFKILRGO-DHC 319
QY 321 GIAS 324
DB 320 GIES 323

Search completed: January 10, 2006, 09:49:50
Job time : 51 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 09:35:49 ; Search time 113 Seconds
(without alignments)
1223.907 Million cell updates/sec

Title: US-10-646-470-1

Perfect score: 1806
Sequence: 1 MKRLVCVLLVCSAVALHKL.....MARNGKHGIAFPSPYPEI 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pdp:*
2: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pdp:*
3: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pdp:*
4: /cgn2_6/ptodata/1/pubppa/US10A_PUBCOMB.pdp:*
5: /cgn2_6/ptodata/1/pubppa/US10B_PUBCOMB.pdp:*
6: /cgn2_6/ptodata/1/pubppa/US11_PUBCOMB.pdp:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1806	100.0	331	4	US-10-099-275-2 Sequence 2, Appl1
2	1800	99.7	331	4	US-10-273-577-6 Sequence 6, Appl1
3	1800	99.7	331	5	US-10-973-858-20 Sequence 20, Appl1
4	1800	99.7	331	5	US-10-287-436A-412 Sequence 412, Appl1
5	1800	99.7	331	5	US-10-287-436A-1109 Sequence 1109, Appl1
6	1797	99.5	331	5	US-10-635-398-96 Sequence 96, Appl1
7	1794	99.3	331	4	US-10-273-577-1 Sequence 1, Appl1
8	1792	99.2	331	5	US-10-809-816A-4 Sequence 4, Appl1
9	1791	99.2	331	3	US-09-953-956-8 Sequence 8, Appl1
10	1791	99.2	331	4	US-10-114-464-8 Sequence 8, Appl1
11	1791	99.2	331	5	US-10-726-645-8 Sequence 8, Appl1
12	1790	99.1	331	4	US-10-273-577-7 Sequence 7, Appl1
13	1784	98.8	331	4	US-10-273-577-8 Sequence 8, Appl1
14	1784	98.8	331	4	US-10-318-584-4 Sequence 4, Appl1
15	1740	96.3	331	5	US-10-809-816A-2 Sequence 2, Appl1
16	1700.5	94.2	330	5	US-10-010-577-2 Sequence 2, Appl1
17	1700.5	94.2	330	5	US-10-894-104-2 Sequence 2, Appl1
18	1700.5	94.2	330	5	US-10-809-816A-3 Sequence 3, Appl1
19	1641	90.9	332	5	US-10-809-816A-7 Sequence 7, Appl1
20	1599	88.5	331	5	US-10-010-580-2 Sequence 2, Appl1
21	1599	88.5	331	5	US-10-894-106-2 Sequence 2, Appl1
22	1599	88.5	331	5	US-10-809-816A-5 Sequence 5, Appl1
23	1578	87.4	331	3	US-09-990-064-2 Sequence 2, Appl1
24	1424	78.8	267	4	US-10-408-765A-1343 Sequence 1343, Appl1
25	1335.5	72.9	330	5	US-10-809-816A-6 Sequence 6, Appl1
26	1188	65.8	217	4	US-10-273-577-2 Sequence 2, Appl1
27	1182	65.4	217	4	US-10-273-577-5 Sequence 5, Appl1

28	1182	65.4	222	5	US-10-635-398-98 Sequence 98, Appl1
29	1178	65.2	217	4	US-10-273-577-3 Sequence 3, Appl1
30	1172	64.9	217	4	US-10-273-577-4 Sequence 4, Appl1
31	952.5	52.7	329	3	US-09-953-956-7 Sequence 7, Appl1
32	952.5	52.7	329	4	US-10-114-464-7 Sequence 7, Appl1
33	952.5	52.7	329	4	US-10-318-584-3 Sequence 3, Appl1
34	952.5	52.7	329	5	US-10-726-645-7 Sequence 7, Appl1
35	952	52.7	330	5	US-10-929-918A-32 Sequence 32, Appl1
36	951.5	52.7	329	4	US-09-953-956-2 Sequence 2, Appl1
37	951.5	52.7	329	4	US-10-114-464-2 Sequence 2, Appl1
38	951.5	52.7	329	4	US-10-318-584-2 Sequence 2, Appl1
39	951.5	52.7	329	5	US-10-723-860-1211 Sequence 1211, Appl1
40	951.5	52.7	329	5	US-10-726-645-2 Sequence 2, Appl1
41	951.5	52.7	329	5	US-10-973-858-18 Sequence 18, Appl1
42	951.5	52.7	329	5	US-10-756-149-5007 Sequence 5007, Appl1
43	951.5	52.7	361	3	US-09-925-300-1693 Sequence 1693, Appl1
44	868	48.1	291	5	US-10-836-897-4 Sequence 4, Appl1
45	865	47.9	334	3	US-09-784-641-4 Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-10-099-275-2
; Sequence 2, Application US/10099275
; Publication No. US20020187499A1
; GENERAL INFORMATION:
; APPLICANT: Schneider, Patrick
; APPLICANT: Yamamoto, Karen K.
; APPLICANT: French, Cynthia K.
; APPLICANT: Reprogen, Inc.
; TITLE OF INVENTION: Use of Cathepsin S in the Diagnosis and Treatment of
; FILE REFERENCE: 018002-001310US
; CURRENT FILING DATE: 2002-04-13
; PRIOR APPLICATION NUMBER: US/10/099,275
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/088,017
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-099-275-2

Query Match 100.0%; Score 1806; DB 4; Length 331;
Beet Local Similarity 100.0%; Pred. No. 1.4e-166; Indels 0; Gaps 0;
Matches 331; Conservative

QY 1 MKRLVCVLLVCSAVALHKLPTLDHMHLMKTYGKYKKNBEAVRLIWEKNLKFVM 60
Db 1 MKRLVCVLLVCSAVALHKLPTLDHMHLMKTYGKYKKNBEAVRLIWEKNLKFVM 60

QY 1 LHNLESHSGMHSYDLSGNHLDGMTSEBVMSLSLRVSQWQRNTTYSNPKRIIPDSVD 120
Db 1 LHNLESHSGMHSYDLSGNHLDGMTSEBVMSLSLRVSQWQRNTTYSNPKRIIPDSVD 120

QY 61 LHNLESHSGMHSYDLSGNHLDGMTSEBVMSLSLRVSQWQRNTTYSNPKRIIPDSVD 120
Db 61 LHNLESHSGMHSYDLSGNHLDGMTSEBVMSLSLRVSQWQRNTTYSNPKRIIPDSVD 120

QY 121 WREKCVTEVKKYQSGCGAFAVAAGALBAOLKLTGTLVLSAONLVDCSTEXTGKNGCC 180
Db 121 WREKCVTEVKKYQSGCGAFAVAAGALBAOLKLTGTLVLSAONLVDCSTEXTGKNGCC 180

QY 181 NGGFMTTAFQYITIDNKGIDSDASYPKAMDLCQYDSRYRATCSKYTELPYGRBDVJKE 240
Db 181 NGGFMTTAFQYITIDNKGIDSDASYPKAMDLCQYDSRYRATCSKYTELPYGRBDVJKE 240

QY 241 AVANKGPVSVGVDAHPSFPLYRSGVYVEPCTQVNVHGVTVVGGDINGEXYWLVRKSW 300
Db 241 AVANKGPVSVGVDAHPSFPLYRSGVYVEPCTQVNVHGVTVVGGDINGEXYWLVRKSW 300

Qy 301 GHNGEEGYIRMARNGKHGCIASFPSPYPEI 331
|||
Db 301 GHNGEEGYIRMARNGKHGCIASFPSPYPEI 331

```

RESULT 2
US-10-273-577-6
; Sequence 6, Application US/10273577
; Publication No. US2003014371A1
; GENERAL INFORMATION:
; APPLICANT: Lamers, Marieke B.
; APPLICANT: Williams, David H.
; TITLE OF INVENTION: Crystal Structure of a Mutant of Cathepsin S Enzyme
; FILE REFERENCE: 1/18-0202P
; CURRENT APPLICATION NUMBER: US/10/273,577
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: 60/330,191
; PRIOR FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 331
; TYPE: PRT
; ORGANISM: human
; US-10-273-577-6

```

Query Match	99.7%	Score 1800;	DB 4;	Length 331;
Best Local Similarity	99.7%	Pred. No. 5,4e-166;		
Matches 330; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

Qy	1	MKRLVCTLLVCSAAVLAOLHKDPTLIDHHMLMKTKYOKYKKEAEVRLIWEKONLKFVM	60
Db	1	MKRLVCTLLVCSAAVLAOLHKDPTLIDHHMLMKTKYOKYKKEAEVRLIWEKONLKFVM	60
Qy	61	LHNLESHMGMSHYDGLGNHNLGDMTSEVNSLSLAVPSOMORNIITYKSNPNRLPDSVD	120
Db	61	LHNLESHMGMSHYDGLGNHNLGDMTSEVNSLSLAVPSOMORNIITYKSNPNRLPDSVD	120
Qy	121	WREKGCVTBEVKYQGSGCAGMAFSVAVALBAOLKLTKGLVLSLAONLVDCKSTERYGNKGC	180
Db	121	WREKGCVTBEVKYQGSGCAGMAFSVAVALBAOLKLTKGLVLSLAONLVDCKSTERYGNKGC	180
Qy	181	NGGFMPTTAFQYIINDNKGIDSDASYPYKAMDLCQYDSTKRAATCSKYTELPYGRBDYLKE	240
Db	181	NGGFMPTTAFQYIINDNKGIDSDASYPYKAMDLCQYDSTKRAATCSKYTELPYGRBDYLKE	240
Qy	241	AVANAGPVSVCVDAAHPSFFFLYRSGYTYEBSCTQNNVHGVLVGYGDLNGKEKFWLYKNSW	300
Db	241	AVANAGPVSVCVDAAHPSFFFLYRSGYTYEBSCTQNNVHGVLVGYGDLNGKEKFWLYKNSW	300
Qy	301	GHNPGEBGYIRMAKRNKGNHCGIASPSPSYBI	331
Db	301	GHNPGEBGYIRMAKRNKGNHCGIASPSPSYBI	331

```

RESULT 3
US-10-973-658-20
Sequence 20, Application US/10973858
Publication No. US20050176030A1
GENERAL INFORMATION:
APPLICANT: Gan, Li
APPLICANT: Gonzalez-Zulueeta, Mirella
APPLICANT: Ye, Shiming
APPLICANT: Uferer, Roman
TITLE OF INVENTION: Regulated Nucleic Acids in Pathogenesis of Alzheimer's Disease
FILE REFERENCE: AGYT-047
CURRENT APPLICATION NUMBER: US/10/973,858
CURRENT FILING DATE: 2004-10-25
PRIOR APPLICATION NUMBER: 60/515,562
PRIOR FILING DATE: 2003-10-28
NUMBER OF SEQ ID NOS: 48
SOFTWARE: PatentIn version 3.2

```

```

; SEQ ID NO 20
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-973-858-20

```

```

Query Match      99.7%; Score 1800; DB 5; Length 331;
Best Local Similarity 99.7%; Pred. No. 5.4e-166;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0

```

QY	1	MKRLVCLVLCSSAAVAOLAKHDPTLHHNHLAKTKYQKQYKXNBEAVRLIIMKXNLKFLVM	60
Db	1	MKRLVCLVLCSSAAVAOLAKHDPTLHHNHLAKTKYQKQYKXNBEAVRLIIMKXNLKFLVM	60
QY	61	LHNLEHSGMHSYDIGNMHLGDMTSEBVNSLMSLRLVPSCQORNTYYKSNPNFLDPDSVD	120
Db	61	LHNLEHSGMHSYDIGNMHLGDMTSEBVNSLMSLRLVPSCQORNTYYKSNPNFLDPDSVD	120
QY	121	WREKGCYVEVKYQGSQCGACMAFSAAGALAEOLKLTGKVLVLSAQNVLVQCSCTEYGNKGC	180
Db	121	WREKGCYVEVKYQGSQCGACMAFSAAGALAEOLKLTGKVLVLSAQNVLVQCSCTEYGNKGC	180
QY	181	NGGFMTTAPQYIINDKGGIDSDASYYKKAMDLKQYDSKTRAAATCSKYTELPYGRBDYLKE	240
Db	181	NGGFMTTAPQYIINDKGGIDSDASYYKKAMDLKQYDSKTRAAATCSKYTELPYGRBDYLKE	240
QY	241	AVANNGPVSVGVDARHPSFLYLRSGYVEPSCQOVNVNHGVLVVGYGDLNGKEVTLVNSW	300
Db	241	AVANNGPVSVGVDARHPSFLYLRSGYVEPSCQOVNVNHGVLVVGYGDLNGKEVTLVNSW	300
QY	301	GHNFGEEGYIMARKNGNHCGLASPSPSYEI	331
Db	301	GHNFGEEGYIMARKNGNHCGLASPSPSYEI	331

```

4 RESULT 4
5 US-10-287-436A-412
6 Sequence 412 Application US/10287436A
7 Publication No US20050202421A1
8 GENERAL INFORMATION:
9 APPLICANT: CHILDREN'S HOSPITAL MEDICAL CENTER
10 TITLE OF INVENTION: METHOD FOR DIAGNOSIS AND
11 TITLE OF INVENTION: RHEUMATOID ARTHRITIS
12 FILE REFERENCE: 10872.514696
13 CURRENT APPLICATION NUMBER: US/10/287,436A
14 PRIOR FILING DATE: 2002-10-31
15 PRIOR FILING DATE: 2001-10-31
16 NUMBER OF SEQ ID NOS: 1446
17 SOFTWARE: FASTSEQ for Windows Version 4.0
18 SEQ ID NO 412
19 LENGTH: 331
20 TYPE: PRT
21 ORGANISM: Homo sapiens
22 US-10-287-436A-412

```

Query Match	99.7%	Score 1800;	DB 5;	Length 331;
Best Local Similarity	99.7%	Pred. No. 5.4e-166;		
Matches 330; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

QY 1 MRILVCVLLVCSAAVAOLHKDPTLDHMHMLAKKTKGQVKEKNEEAVRRLIEMKONLKFVM 60

Db 1 MRILVCVLLVCSAAVAOLHKDPTLDHMHMLAKKTKGQVKEKNEEAVRRLIEMKONLKFVM 60

QY 61 LHLNLSHSMGMSYDLDGMNHLGDMTSEEVLSLSSLSLPSQWQRNITYKSPNRLILPDSVD 120

Db 61 LHLNLSHSMGMSYDLDGMNHLGDMTSEEVLSLSSLSLPSQWQRNITYKSPNRLILPDSVD 120

QY 121 WREKCVTEVKKVQSGCGACMAFSAVGALTEAOLKLTKGKLVSLSAQNLVDCSTERYGNKGC 180

Db 121 WREKCVTEVKKVQSGCGACMAFSAVGALTEAOLKLTKGKLVSLSAQNLVDCSTERYGNKGC 180

QY 161 NGGFMTTAPQVYITDNGKIDSDASYPYKAMDLCQVDSKTRATCSKYTELPIYGRVDYLKE 240

```

Db      181 NGCFMTTAFQYIIDKIGDSDASYKAMDQCKQYDSKTRATCSKYTELPYGRDVLKE 240
QY      241 AVANKGPVSVGVDAHPSFFLYRSGVYYPEPSCQVNVNMGVLVVGDLNGKEYMLVKNSW 300
Db      241 AVANKGPVSVGVDAHPSFFLYRSGVYYPEPSCQVNVNMGVLVVGDLNGKEYMLVKNSW 300
QY      301 GHNFGEEGYIRMARKNKGHCIGIASPSPYPEI 331
Db      301 GHNFGEEGYIRMARKNKGHCIGIASPSPYPEI 331

```

RESULT 5

```

US-10-287-436A-1109
; Sequence 1109, Application US/10287436A
; Publication No. US20050202421A1
; GENERAL INFORMATION:
; APPLICANT: CHILDREN'S HOSPITAL MEDICAL CENTER
; TITLE OF INVENTION: METHOD FOR DIAGNOSIS AND TREATMENT OF
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS
; FILE REFERENCE: 10872.514596
; CURRENT APPLICATION NUMBER: US/10/287,436A
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/336,220
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 1446
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1109
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-287-436A-1109

```

```

Query Match      99.7%; Score 1800; DB 5; Length 331;
Best Local Similarity 99.7%; Pred. No. 5.4e-166;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 MKRLVCVLLVCSAQAOLKDPDLDHMHLMKKTGYKQYKENEAEVRLIWEKLLKFFM 60
Db      1 MKRLVCVLLVCSAQAOLKDPDLDHMHLMKKTGYKQYKENEAEVRLIWEKLLKFFM 60
QY      61 LHNLEHSGMHSYDYGNNHLDGMTSEBVMSLSLRVPSQWQRNITYKSNPRLIPDSVD 120
Db      61 LHNLEHSGMHSYDYGNNHLDGMTSEBVMSLSLRVPSQWQRNITYKSNPRLIPDSVD 120
QY      121 WREKGCCTEVKYQSGCGAFAVGALEAQLKLTGKLVLSAQNLDVDCSTEEKYGNKGC 180
Db      121 WREKGCCTEVKYQSGCGAFAVGALEAQLKLTGKLVLSAQNLDVDCSTEEKYGNKGC 180
QY      181 NGCFMTTAFQYIIDKIGDSDASYKAMDLCQYDSKTRATCSKYTELPYGRDVLKE 240
Db      181 NGCFMTTAFQYIIDKIGDSDASYKAMDLCQYDSKTRATCSKYTELPYGRDVLKE 240
QY      241 AVANKGPVSVGVDAHPSFFLYRSGVYYPEPSCQVNVNMGVLVVGDLNGKEYMLVKNSW 300
Db      241 AVANKGPVSVGVDAHPSFFLYRSGVYYPEPSCQVNVNMGVLVVGDLNGKEYMLVKNSW 300
QY      301 GHNFGEEGYIRMARKNKGHCIGIASPSPYPEI 331
Db      301 GHNFGEEGYIRMARKNKGHCIGIASPSPYPEI 331

```

RESULT 6

```

US-10-635-398-96
; Sequence 96, Application US/10635398
; Publication No. US20050037957A1
; GENERAL INFORMATION:
; APPLICANT: David Anderson
; APPLICANT: Constance Berghs
; APPLICANT: Elina Catterton
; APPLICANT: Shlomit Edinger
; APPLICANT: Linda Gorman
; APPLICANT: Xiaojia (Sabha) Guo

```

```

; APPLICANT: John Hermann
; APPLICANT: Ramesh Kekuda
; APPLICANT: Li Li
; APPLICANT: Daniel Rieger
; APPLICANT: Wei Zhong
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-593 C
; CURRENT APPLICATION NUMBER: US/10/635,398
; CURRENT FILING DATE: 2003-08-06
; PRIOR APPLICATION NUMBER: 10/160,619
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: 60/295,661
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/359,122
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/296,404
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/359,035
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 60/299,949
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: 60/295,607
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/359,964
; PRIOR FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: 60/296,418
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/341,562
; PRIOR FILING DATE: 2001-12-14
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: Cursedlist version 0.1
; SEQ ID NO 96
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-635-398-96

```

```

Query Match      99.5%; Score 1797; DB 5; Length 331;
Best Local Similarity 99.4%; Pred. No. 1.1e-165;
Matches 329; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 MKRLVCVLLVCSAQAOLKDPDLDHMHLMKKTGYKQYKENEAEVRLIWEKLLKFFM 60
Db      1 MKRLVCVLLVCSAQAOLKDPDLDHMHLMKKTGYKQYKENEAEVRLIWEKLLKFFM 60
QY      61 LHNLEHSGMHSYDYGNNHLDGMTSEBVMSLSLRVPSQWQRNITYKSNPRLIPDSVD 120
Db      61 LHNLEHSGMHSYDYGNNHLDGMTSEBVMSLSLRVPSQWQRNITYKSNPRLIPDSVD 120
QY      121 WREKGCCTEVKYQSGCGAFAVGALEAQLKLTGKLVLSAQNLDVDCSTEEKYGNKGC 180
Db      121 WREKGCCTEVKYQSGCGAFAVGALEAQLKLTGKLVLSAQNLDVDCSTEEKYGNKGC 180
QY      181 NGCFMTTAFQYIIDKIGDSDASYKAMDLCQYDSKTRATCSKYTELPYGRDVLKE 240
Db      181 NGCFMTTAFQYIIDKIGDSDASYKAMDLCQYDSKTRATCSKYTELPYGRDVLKE 240
QY      241 AVANKGPVSVGVDAHPSFFLYRSGVYYPEPSCQVNVNMGVLVVGDLNGKEYMLVKNSW 300
Db      241 AVANKGPVSVGVDAHPSFFLYRSGVYYPEPSCQVNVNMGVLVVGDLNGKEYMLVKNSW 300
QY      301 GHNFGEEGYIRMARKNKGHCIGIASPSPYPEI 331
Db      301 GHNFGEEGYIRMARKNKGHCIGIASPSPYPEI 331

```

RESULT 7

```

US-10-273-577-1
; Sequence 1, Application US/10273577
; Publication No. US2003014371A1
; GENERAL INFORMATION:
; APPLICANT: Lamers, Marijke B.

```

APPLICANT: Williams, David H.
; TITLE OF INVENTION: Crystal Structure of a Mutant of Cathepsin S Enzyme
; FILE REFERENCE: 1718-0202P
; CURRENT APPLICATION NUMBER: US/10/273,577
; PRIOR FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: 60/330,191
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 331
; TYPE: PRT
; ORGANISM: human
US-10-273-577-1

Query Match 99.3%; Score 1794; DB 4; Length 331;
Best Local Similarity 99.4%; Pred. No. 2,1e-165;
Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVCSAVALQHLKDPDLDHMHLMKKTGYKQYKKNBEAVRLLIWEKNLKFVM 60
DB 1 MKRLVCVLLVCSAVALQHLKDPDLDHMHLMKKTGYKQYKKNBEAVRLLIWEKNLKFVM 60
QY 61 LHNLEHSMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMQRNTTYSNPNRLLPDSVD 120
DB 61 LHNLEHSMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMQRNTTYSNPNRLLPDSVD 120
QY 121 WREKCVTEVKYQSGCGMAFSAVGALBAOLKLTGKLVSLSAONLVDCTEKGNGKC 180
DB 121 WREKCVTEVKYQSGCGMAFSAVGALBAOLKLTGKLVSLSAONLVDCTEKGNGKC 180
QY 181 NGGFMTTAFQYIINDKIGDSDASYPKAMDLCQYDSKRYRATCSKYTELPYGRDVLKE 240
DB 181 NGGFMTTAFQYIINDKIGDSDASYPKAMDLCQYDSKRYRATCSKYTELPYGRDVLKE 240
QY 241 AVANGPVSVGVDARHPSFPLYRSGVYEPSCQVNVHGVLVVGYGDLNGKEXYMLVKNSW 300
DB 241 AVANGPVSVGVDARHPSFPLYRSGVYEPSCQVNVHGVLVVGYGDLNGKEXYMLVKNSW 300
QY 301 GHNFGEBGYIMARKGNHCGIASPSPYPEI 331
DB 301 GHNFGEBGYIMARKGNHCGIASPSPYPEI 331

RESULT 8
US-10-809-816A-4
; Sequence 4, Application US/10809816A
; Publication No. US20050214774A1
; GENERAL INFORMATION:
; APPLICANT: Li, Shyr-Jiann et al.
; TITLE OF INVENTION: ISOLATED MONKEY CATHEPSIN S PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING MONKEY CATHEPSIN S PROTEINS,
; FILE REFERENCE: C1001507
; CURRENT APPLICATION NUMBER: US/10/809,816A
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-809-816A-4

Query Match 99.2%; Score 1792; DB 5; Length 331;
Best Local Similarity 99.4%; Pred. No. 3.3e-165;
Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVCSAVALQHLKDPDLDHMHLMKKTGYKQYKKNBEAVRLLIWEKNLKFVM 60
DB 1 MKRLVCVLLVCSAVALQHLKDPDLDHMHLMKKTGYKQYKKNBEAVRLLIWEKNLKFVM 60
QY 61 LHNLEHSMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMQRNTTYSNPNRLLPDSVD 120

DB 61 LHNLEHSMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMQRNTTYSNPNRLLPDSVD 120
QY 121 WREKCVTEVKYQSGCGMAFSAVGALBAOLKLTGKLVSLSAONLVDCTEKGNGKC 180
DB 121 WREKCVTEVKYQSGCGMAFSAVGALBAOLKLTGKLVSLSAONLVDCTEKGNGKC 180
QY 181 NGGFMTTAFQYIINDKIGDSDASYPKAMDLCQYDSKRYRATCSKYTELPYGRDVLKE 240
DB 181 NGGFMTTAFQYIINDKIGDSDASYPKAMDLCQYDSKRYRATCSKYTELPYGRDVLKE 240
QY 241 AVANGPVSVGVDARHPSFPLYRSGVYEPSCQVNVHGVLVVGYGDLNGKEXYMLVKNSW 300
DB 241 AVANGPVSVGVDARHPSFPLYRSGVYEPSCQVNVHGVLVVGYGDLNGKEXYMLVKNSW 300
QY 301 GHNFGEBGYIMARKGNHCGIASPSPYPEI 331
DB 301 GHNFGEBGYIMARKGNHCGIASPSPYPEI 331

RESULT 9
US-09-953-956-8
; Sequence 8, Application US/09953956
; Patent No. US20020072107A1
; GENERAL INFORMATION:
; APPLICANT: Haselings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PFI07D2D1
; CURRENT APPLICATION NUMBER: US/09/953,956
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: 09/219,441
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-953-956-8

Query Match 99.2%; Score 1791; DB 3; Length 331;
Best Local Similarity 99.1%; Pred. No. 4.1e-165;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVCSAVALQHLKDPDLDHMHLMKKTGYKQYKKNBEAVRLLIWEKNLKFVM 60
DB 1 MKRLVCVLLVCSAVALQHLKDPDLDHMHLMKKTGYKQYKKNBEAVRLLIWEKNLKFVM 60
QY 61 LHNLEHSMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMQRNTTYSNPNRLLPDSVD 120
DB 61 LHNLEHSMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMQRNTTYSNPNRLLPDSVD 120
QY 121 WREKCVTEVKYQSGCGMAFSAVGALBAOLKLTGKLVSLSAONLVDCTEKGNGKC 180
DB 121 WREKCVTEVKYQSGCGMAFSAVGALBAOLKLTGKLVSLSAONLVDCTEKGNGKC 180
QY 181 NGGFMTTAFQYIINDKIGDSDASYPKAMDLCQYDSKRYRATCSKYTELPYGRDVLKE 240
DB 181 NGGFMTTAFQYIINDKIGDSDASYPKAMDLCQYDSKRYRATCSKYTELPYGRDVLKE 240
QY 241 AVANGPVSVGVDARHPSFPLYRSGVYEPSCQVNVHGVLVVGYGDLNGKEXYMLVKNSW 300
DB 241 AVANGPVSVGVDARHPSFPLYRSGVYEPSCQVNVHGVLVVGYGDLNGKEXYMLVKNSW 300
QY 301 GHNFGEBGYIMARKGNHCGIASPSPYPEI 331
DB 301 GHNFGEBGYIMARKGNHCGIASPSPYPEI 331

RESULT 10
US-10-114-464-8
; Sequence 8, Application US/10114464
; Publication No. US20020142448A1

GENERAL INFORMATION:
APPLICANT: Hastings, et al.
TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: PFI07D5
CURRENT APPLICATION NUMBER: US/10/114,464
CURRENT FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 08/553,125
PRIOR FILING DATE: 1995-11-07
PRIOR APPLICATION NUMBER: 08/208,007
PRIOR FILING DATE: 1994-03-08
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 331
TYPE: PRT
ORGANISM: Homo sapiens
US-10-114-464-8

Query Match 99.2%; Score 1791; DB 4; Length 331;
Best Local Similarity 99.1%; Pred. No. 4,1e-165;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVSSSAVAOLHDKDPTLDHWHLMKKTGYKQYKEKNEBAVRLLIWEKXKLFVM 60
DB 1 MKRLVCVLLVSSSAVAOLHDKDPTLDHWHLMKKTGYKQYKEKNEBAVRLLIWEKXKLFVM 60
QY 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQWQRNTTYSNPRLLIPDSVD 120
DB 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQWQRNTTYSNPRLLIPDSVD 120
QY 121 WREKCVTEVVKYQSCGCMAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTEXKGNKC 180
DB 121 WREKCVTEVVKYQSCGCMAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTEXKGNKC 180
QY 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
DB 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
QY 241 AVANGPVSQVVDARHPSFFLYRSGVYEPSCQNVNNGVLVVGDLNGKEXYMLVKNWS 300
DB 241 AVANGPVSQVVDARHPSFFLYRSGVYEPSCQNVNNGVLVVGDLNGKEXYMLVKNWS 300
QY 301 GHNFGEEGYIRMAKNKGNHCGIASFPSPYPEI 331
DB 301 GHNFGEEGYIRMAKNKGNHCGIASFPSPYPEI 331

RESULT 11
US-10-726-645-8
Sequence 8, Application US/10726645
Publication No. US20050019889A1
GENERAL INFORMATION:
APPLICANT: Hastings, et al.
TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: PFI07D5
CURRENT APPLICATION NUMBER: US/10/726,645
CURRENT FILING DATE: 2003-12-04
PRIOR APPLICATION NUMBER: US/10/114,464
PRIOR FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 08/553,125
PRIOR FILING DATE: 1995-11-07
PRIOR APPLICATION NUMBER: 08/208,007
PRIOR FILING DATE: 1994-03-08
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 331
TYPE: PRT
ORGANISM: Homo sapiens
US-10-726-645-8

Query Match 99.2%; Score 1791; DB 5; Length 331;
Best Local Similarity 99.1%; Pred. No. 4,1e-165;

Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVSSSAVAOLHDKDPTLDHWHLMKKTGYKQYKEKNEBAVRLLIWEKXKLFVM 60
DB 1 MKRLVCVLLVSSSAVAOLHDKDPTLDHWHLMKKTGYKQYKEKNEBAVRLLIWEKXKLFVM 60
QY 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQWQRNTTYSNPRLLIPDSVD 120
DB 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQWQRNTTYSNPRLLIPDSVD 120
QY 121 WREKCVTEVVKYQSCGCMAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTEXKGNKC 180
DB 121 WREKCVTEVVKYQSCGCMAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTEXKGNKC 180
QY 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
DB 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
QY 241 AVANGPVSQVVDARHPSFFLYRSGVYEPSCQNVNNGVLVVGDLNGKEXYMLVKNWS 300
DB 241 AVANGPVSQVVDARHPSFFLYRSGVYEPSCQNVNNGVLVVGDLNGKEXYMLVKNWS 300
QY 301 GHNFGEEGYIRMAKNKGNHCGIASFPSPYPEI 331
DB 301 GHNFGEEGYIRMAKNKGNHCGIASFPSPYPEI 331

RESULT 12
US-10-273-577-7
Sequence 7, Application US/10273577
Publication No. US20030143714A1
GENERAL INFORMATION:
APPLICANT: Williams, David H.
TITLE OF INVENTION: Crystall Structure of a Mutant of Cathepsin S Enzyme
FILE REFERENCE: 1718-0202P
CURRENT APPLICATION NUMBER: US/10/273,577
CURRENT FILING DATE: 2003-02-14
PRIOR APPLICATION NUMBER: 60/330,191
PRIOR FILING DATE: 2001-10-19
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 331
TYPE: PRT
ORGANISM: human
US-10-273-577-7

Query Match 99.1%; Score 1790; DB 4; Length 331;
Best Local Similarity 99.4%; Pred. No. 5,1e-165;
Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVSSSAVAOLHDKDPTLDHWHLMKKTGYKQYKEKNEBAVRLLIWEKXKLFVM 60
DB 1 MKRLVCVLLVSSSAVAOLHDKDPTLDHWHLMKKTGYKQYKEKNEBAVRLLIWEKXKLFVM 60
QY 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQWQRNTTYSNPRLLIPDSVD 120
DB 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQWQRNTTYSNPRLLIPDSVD 120
QY 121 WREKCVTEVVKYQSCGCMAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTEXKGNKC 180
DB 121 WREKCVTEVVKYQSCGCMAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTEXKGNKC 180
QY 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
DB 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
QY 241 AVANGPVSQVVDARHPSFFLYRSGVYEPSCQNVNNGVLVVGDLNGKEXYMLVKNWS 300
DB 241 AVANGPVSQVVDARHPSFFLYRSGVYEPSCQNVNNGVLVVGDLNGKEXYMLVKNWS 300
QY 301 GHNFGEEGYIRMAKNKGNHCGIASFPSPYPEI 331
DB 301 GHNFGEEGYIRMAKNKGNHCGIASFPSPYPEI 331

Db 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331

RESULT 13

US-10-273-577-8
Sequence 8, Application US/10273577
Publication No. US20030143714A1
GENERAL INFORMATION:
APPLICANT: Lamers, Marieke B.
APPLICANT: Williams, David H.
TITLE OF INVENTION: Crystal Structure of a Mutant of Cathepsin S Enzyme
FILE REFERENCE: 1718-0202P
CURRENT APPLICATION NUMBER: US/10/273,577
CURRENT FILING DATE: 2003-02-14
PRIOR APPLICATION NUMBER: 60/330,191
PRIORITY FILING DATE: 2001-10-19
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 8
LENGTH: 331
TYPE: PRT
ORGANISM: human
US-10-273-577-8

Query Match 98.8%; Score 1784; DB 4; Length 331;
Best Local Similarity 99.1%; Pred. No. 2e-164;
Matches 328; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVCSAVALQKHDPDLDHNLHMKTKYQKQYKKEKEAEAVRLLMEKLLKFTVM 60
Db 1 MKRLVCVLLVCSAVALQKHDPDLDHNLHMKTKYQKQYKKEKEAEAVRLLMEKLLKFTVM 60
QY 61 LHNLESHGMSYDGLGNHLDGMTSEEVMSLSLRVPSQWRNITYSNPRLIPDSVD 120
Db 61 LHNLESHGMSYDGLGNHLDGMTSEEVMSLSLRVPSQWRNITYSNPRLIPDSVD 120
QY 121 WREKCVTEVYKQSCGACMAFSAVGALEAOLKLTGKLVLSAQNLDVDCSTKYGKNGCC 180
Db 121 WREKCVTEVYKQSCGACMAFSAVGALEAOLKLTGKLVLSAQNLDVDCSTKYGKNGCC 180
QY 181 NGFFMTTFQYIINDKGIIDSDASYPKAMDLCQYDSKRYRAATCKYTELPYGRBDVJKE 240
Db 181 NGFFMTTFQYIINDKGIIDSDASYPKAMDLCQYDSKRYRAATCKYTELPYGRBDVJKE 240
QY 241 AVANGPVSVCVDARHPSPFLYRSGVYVEPSCQNVNNGVLVVGVDLNGKEXYMLVKXSM 300
Db 241 AVANGPVSVCVDARHPSPFLYRSGVYVEPSCQNVNNGVLVVGVDLNGKEXYMLVKXSM 300
QY 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331
Db 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331

RESULT 14

US-10-318-584-4
Sequence 4, Application US/10318584
Publication No. US20030175937A1
GENERAL INFORMATION:
APPLICANT: Bromme, Dieter
APPLICANT: Okamoto, Kathleen
TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarradero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/318,584
FILING DATE: 13-Dec-2002
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-60261-1/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-318-584-4

Query Match 98.8%; Score 1784; DB 4; Length 331;
Best Local Similarity 99.1%; Pred. No. 2e-164;
Matches 328; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVCSAVALQKHDPDLDHNLHMKTKYQKQYKKEKEAEAVRLLMEKLLKFTVM 60
Db 1 MKRLVCVLLVCSAVALQKHDPDLDHNLHMKTKYQKQYKKEKEAEAVRLLMEKLLKFTVM 60
QY 61 LHNLESHGMSYDGLGNHLDGMTSEEVMSLSLRVPSQWRNITYSNPRLIPDSVD 120
Db 61 LHNLESHGMSYDGLGNHLDGMTSEEVMSLSLRVPSQWRNITYSNPRLIPDSVD 120
QY 121 WREKCVTEVYKQSCGACMAFSAVGALEAOLKLTGKLVLSAQNLDVDCSTKYGKNGCC 180
Db 121 WREKCVTEVYKQSCGACMAFSAVGALEAOLKLTGKLVLSAQNLDVDCSTKYGKNGCC 180
QY 181 NGFFMTTFQYIINDKGIIDSDASYPKAMDLCQYDSKRYRAATCKYTELPYGRBDVJKE 240
Db 181 NGFFMTTFQYIINDKGIIDSDASYPKAMDLCQYDSKRYRAATCKYTELPYGRBDVJKE 240
QY 241 AVANGPVSVCVDARHPSPFLYRSGVYVEPSCQNVNNGVLVVGVDLNGKEXYMLVKXSM 300
Db 241 AVANGPVSVCVDARHPSPFLYRSGVYVEPSCQNVNNGVLVVGVDLNGKEXYMLVKXSM 300
QY 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331
Db 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331

RESULT 15

US-10-809-816A-2
Sequence 2, Application US/10809816A
Publication No. US20050214774A1
GENERAL INFORMATION:
APPLICANT: Li, Shyr-Jiann et al.
TITLE OF INVENTION: ISOLATED MONKEY CATHEPSIN S PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING MONKEY CATHEPSIN S PROTEINS,
FILE REFERENCE: CLO01507
CURRENT APPLICATION NUMBER: US/10/809,816A
CURRENT FILING DATE: 2004-03-26
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FaecSeq for Windows Version 4.0
SEQ ID NO: 2
LENGTH: 331
TYPE: PRT
ORGANISM: Cynomolgus monkey
US-10-809-816A-2

Query Match 96.3%; Score 1740; DB 5; Length 331;
Best Local Similarity 96.7%; Pred. No. 3.7e-160;
Matches 320; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy	1	MKRLVCLLVCSAVALHDKDPTLDHMHLMKTTYGKQYKEKNEBAVRLLIWEKNLKPYM	60
Db	1	MKQLVCLLVCSAVALHDKDPTLDHMHLMKTTYGKQYKEKNEBAVRLLIWEKNLKPYM	60
Qy	61	LHNLESHMGMSYDLGMMHLDGMTSEEVMSLSLRVPSQWQORNTTYKSNPRLLPDSDV	120
Db	61	LHNLESHMGMSYDLGMMHLDGMTSEEVMSLSLRVPSQWQORNTTYKSNANQILPDSVD	120
Qy	121	WREKCVTEVKYQSCGACMAFSAVGALBAQLKLTGKLVSLSAQNLVDCSTEXYGNKGC	180
Db	121	WREKCVTEVKYQSCGACMAFSAVGALBAQLKLTGKLVSLSAQNLVDCSTEXYGNKGC	180
Qy	181	NGGFMTTAPQYIIDNKGIDSDASYPYKAMDLCQYDSKTYRAATCSKYTELPGREDVLKE	240
Db	181	NGGFMTTAPQYIIDNNGIDSDASYPYKATDQCKQYDSKYRAATCSKYTELPGREDVLKE	240
Qy	241	AVANGPVSVGVDARHPSFELYRSGVYYRPSCTQNVNNGVLYVGYGDLNGKEYLVKNSW	300
Db	241	VVANGPVSVGVDARHPSFELYRSGVYYRPSCTQNVNNGVLYVGYGDLNGKEYLVKNSW	300
Qy	301	GHNFGEEGYIRMARNGNHCGLASFPSPYPEI	331
Db	301	GRNFGEEGYIRMARNGNHCGLASFPSPYPEI	331

Search completed: January 10, 2006, 09:48:48
Job time : 114 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 09:32:19 ; Search time 135 Seconds
(without alignments)
1077.292 Million cell updates/sec

Title: US-10-646-470-1
Perfect score: 1806
Sequence: 1 MKRLVCVLVCSSAVALQHK.....NARKNGHGIASFSPYPEI 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_21:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*
- 9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1806	100.0	331	3	AAV59634 Human Cat
2	1800	99.7	331	7	ADD35932 Human cat
3	1800	99.7	331	7	ADP65168 Human cat
4	1800	99.7	331	9	ADY14944 PRO polyp
5	1800	99.7	331	9	ADY14946 PRO polyp
6	1800	99.7	331	9	AEC01587 Human cat
7	1797	99.5	331	7	ADD45416 Human Pro
8	1797	99.5	331	7	ADP56902 Human Pro
9	1797	99.5	331	8	ADJ71695 Human NOV
10	1794	99.3	331	8	ADD35927 Human cat
11	1791	99.2	331	9	ADY19784 PRO polyp
12	1790	99.1	331	7	ADD35933 Human cat
13	1784	98.8	331	7	ADD35934 Human cat
14	1700.5	94.2	330	6	AAO30451 Monkey ca
15	1678	92.9	313	8	ABM83324 Human dia
16	1678	92.9	313	8	ABM83323 Human dia
17	1599	88.5	331	7	ABR56160 Canine ca
18	1578	87.4	331	6	ABG73437 Human cat
19	1424	78.8	267	7	ADP69537 Human hea
20	1299.5	72.0	330	7	ADD45414 Rat Prote
21	1299.5	72.0	330	7	ADP56900 Rat Prote
22	1188	65.8	217	7	ADD35928 Human mat
23	1182	65.4	217	7	ADD35931 Human mat
24	1182	65.4	222	8	ADJ71697 Human NOV

25	1179	65.3	217	9	ADY58885 Human cat
26	1178	65.2	217	7	ADD35929 Human mat
27	1172	64.9	217	7	ADD35930 Human mat
28	953.5	52.8	329	2	AAW77071 Rat Cache
29	952	52.7	330	2	ADZ11834 Canine ca
30	951.5	52.7	329	2	AAW82720 Human obo
31	951.5	52.7	329	2	AAW95599 Prepro-ca
32	951.5	52.7	329	2	AAW41645 Human cat
33	951.5	52.7	329	2	AAW39216 Human cat
34	951.5	52.7	329	2	AAW81000 Human pro
35	951.5	52.7	329	7	ADP60397 Human Pro
36	951.5	52.7	329	7	ADP45861 Human Pro
37	951.5	52.7	329	8	ADP12997 Protein e
38	951.5	52.7	329	8	ADQ18392 Human sof
39	951.5	52.7	329	8	ABM80141 Tumour-as
40	951.5	52.7	329	8	ADP25400 PRO polyp
41	951.5	52.7	329	9	AEC01585 Human cat
42	951.5	52.7	361	3	AAW57115 Human pro
43	951	52.7	329	7	ADP60395 Rat Prote
44	951	52.7	329	7	ADD45859 Rat Prote
45	948.5	52.5	329	3	AAW30506 A monkey

ALIGNMENTS

RESULT 1
AAV59634
ID AAV59634 standard; protein, 331 AA.

AAV59634;
27-MAR-2000 (first entry)

Human Cathepsin S amino acid sequence.

Cathepsin S; human; endometriosis; treatment; diagnose.

Homo sapiens.

W09963115-A2.
09-DEC-1999.
03-JUN-1999; 99WO-US012335.
04-JUN-1998; 98US-0088017P.
(REPR-) REPROGEN INC.
Schneider P, Yamamoto KK, French CK;
WPI; 2000-086986/07.
N-PSDB; AAZ56150.

Use of cathepsin S in the diagnosis and treatment of endometriosis.

Example; Page 15; 60pp; English.

This is the human cathepsin S protein sequence. Detecting levels of the cathepsin S gene product in a sample compared to a control sample can be used as a method of diagnosing endometriosis. The invention also relates to a method for treating endometriosis through the administration of a probe comprising a detectable label and a ligand that specifically binds to cathepsin S gene product to the subject. The endometriotic lesion can be identified in situ by locating bound labelled probe; and the lesion can be excised. The methods are useful to diagnose, monitor the progress of and treat endometriosis in a subject. The methods are also useful for screening for modulators of cathepsin S gene production in endometrial cells. Antisense cathepsin S gene oligonucleotides are useful for the treatment of endometriosis by down-regulating cathepsin S genes

Sequence 331 AA;

Query Match 100.0%; Score 1806; DB 3; Length 331;
 Best Local Similarity 100.0%; Pred. No. 3e-171;
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRLVLCVLLVSSAAVAQLHKDPTLDHMLMKTKYQKQYKKNBAVRRLIWEKXUKFYM 60
 DB 1 MRLVLCVLLVSSAAVAQLHKDPTLDHMLMKTKYQKQYKKNBAVRRLIWEKXUKFYM 60
 QY 61 LNLHLSHMGMSYDYGNNHLDGMTSEEVMSLSLRVPSQKORNTTYSNPNRIIPDSVD 120
 DB 61 LNLHLSHMGMSYDYGNNHLDGMTSEEVMSLSLRVPSQKORNTTYSNPNRIIPDSVD 120
 QY 121 WREKCVTEVKYQSGCGACMAFSAVGALBAOLKLTGKLVLSAQNLDVDCSTERYGNKGC 180
 DB 121 WREKCVTEVKYQSGCGACMAFSAVGALBAOLKLTGKLVLSAQNLDVDCSTERYGNKGC 180
 QY 181 NGGFMTTAFQYTIIDNKGIDSDASYPKAMDLCQYDSKXRAATCSKYTELPYGRBDVLKE 240
 DB 181 NGGFMTTAFQYTIIDNKGIDSDASYPKAMDLCQYDSKXRAATCSKYTELPYGRBDVLKE 240
 QY 241 AVANKGPVSVGVDAHPSFELYSRGVYEPSCQNVNNGVLVVGVDLNGKEVWLKXNSW 300
 DB 241 AVANKGPVSVGVDAHPSFELYSRGVYEPSCQNVNNGVLVVGVDLNGKEVWLKXNSW 300
 QY 301 GHNFGEGYIRMARKNKGNHCGIASPSPYPEI 331
 DB 301 GHNFGEGYIRMARKNKGNHCGIASPSPYPEI 331

RESULT 2
 ADD35932
 ID ADD35932 standard; peptide; 331 AA.
 AC ADD35932;
 XX 15-JAN-2004 (first entry)
 DT 15-JAN-2004 (first entry)
 XX Human cathepsin S mutant #1.
 DE Human cathepsin S; cats inhibitor; human; mutant; mutein.
 KW Cystal; cathepsin S; cats inhibitor; human; mutant; mutein.
 XX Synchetic.
 OS Homo sapiens.
 XX US2003143714-A1.
 PN 31-JUL-2003.
 PD 31-JUL-2003.
 XX 18-OCT-2002; 2002US-00273577.
 PF 18-OCT-2002; 2002US-00273577.
 XX 19-OCT-2001; 2001US-0330191P.
 PR 19-OCT-2001; 2001US-0330191P.
 XX (MEDI-) MEDIVIR UK LTD.
 PA (MEDI-) MEDIVIR UK LTD.
 PI Lamers MBAC, Williams DH, Turkenburg JP, Hubbard RE;
 XX WPI; 2003-829792/77.
 DR WPI; 2003-829792/77.
 XX Crystalline cathepsin S polypeptide free of irreversible inhibitor useful
 PT for identifying and producing potential cats inhibitor.
 XX Claim 2; SEQ ID NO 6; 59pp; English.
 XX The invention relates to a crystalline cathepsin S polypeptide. The
 CC crystalline cats polypeptide is useful for identifying a potential cats
 CC inhibitor molecule. The present sequence represents the amino acid
 CC sequence of a human cathepsin S mutant.
 SQ Sequence 331 AA;
 Query Match 99.7%; Score 1800; DB 7; Length 331;
 Best Local Similarity 99.7%; Pred. No. 1.2e-170;

Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRLVLCVLLVSSAAVAQLHKDPTLDHMLMKTKYQKQYKKNBAVRRLIWEKXUKFYM 60
 DB 1 MRLVLCVLLVSSAAVAQLHKDPTLDHMLMKTKYQKQYKKNBAVRRLIWEKXUKFYM 60
 QY 61 LNLHLSHMGMSYDYGNNHLDGMTSEEVMSLSLRVPSQKORNTTYSNPNRIIPDSVD 120
 DB 61 LNLHLSHMGMSYDYGNNHLDGMTSEEVMSLSLRVPSQKORNTTYSNPNRIIPDSVD 120
 QY 121 WREKCVTEVKYQSGCGACMAFSAVGALBAOLKLTGKLVLSAQNLDVDCSTERYGNKGC 180
 DB 121 WREKCVTEVKYQSGCGACMAFSAVGALBAOLKLTGKLVLSAQNLDVDCSTERYGNKGC 180
 QY 181 NGGFMTTAFQYTIIDNKGIDSDASYPKAMDLCQYDSKXRAATCSKYTELPYGRBDVLKE 240
 DB 181 NGGFMTTAFQYTIIDNKGIDSDASYPKAMDLCQYDSKXRAATCSKYTELPYGRBDVLKE 240
 QY 241 AVANKGPVSVGVDAHPSFELYSRGVYEPSCQNVNNGVLVVGVDLNGKEVWLKXNSW 300
 DB 241 AVANKGPVSVGVDAHPSFELYSRGVYEPSCQNVNNGVLVVGVDLNGKEVWLKXNSW 300
 QY 301 GHNFGEGYIRMARKNKGNHCGIASPSPYPEI 331
 DB 301 GHNFGEGYIRMARKNKGNHCGIASPSPYPEI 331

RESULT 3
 ADP65168
 ID ADP65168 standard; protein; 331 AA.
 AC ADP65168;
 XX 12-AUG-2004 (first entry)
 DT 12-AUG-2004 (first entry)
 XX Human cathepsin S preproprotein.
 DE Human cathepsin S preproprotein.
 KW autoimmune disease; arthritis; gene expression analysis;
 KW rheumatoid arthritis; collagen-induced; immunosuppressive; antirheumatic;
 KW antiarthritic; osteopathic; antigout; antiinflammatory; dermatological;
 KW immunomodulatory; lupus; ankylosing spondylitis; fibrositis;
 KW fibromyalgia; osteoarthritis; gout; juvenile rheumatoid arthritis;
 KW immune; human.
 XX Homo sapiens.
 OS Homo sapiens.
 XX WO2003072827-A1.
 PN WO2003072827-A1.
 PD 04-SEP-2003.
 XX 31-OCT-2002; 2002WO-US035433.
 PF 31-OCT-2002; 2002WO-US035433.
 XX 31-OCT-2001; 2001US-0336220P.
 PR 31-OCT-2001; 2001US-0336220P.
 XX (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
 PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
 PI Hirsch R, Thorton SL;
 XX Hirsch R, Thorton SL;
 DR WPI; 2003-712740/67.
 DR GENBANK; NF_004070.
 XX Diagnosing and analyzing autoimmune disease using gene expression
 PT profiles and microarray technology, useful for diagnosing and treating
 PT rheumatoid arthritis, lupus, fibrositis, osteoarthritis, fibromyalgia and
 PT gout.
 XX Disclosure; Page; 56pp; English.
 XX The invention relates to a novel method for diagnosing and analyzing
 CC autoimmune disease or arthritides. The method comprises obtaining a
 CC patient sample containing mRNA, analyzing gene expression using the mRNA
 CC that results in a gene expression signature of the mRNA, and using that
 CC gene expression signature to diagnose or analyse the autoimmune disease

or arthritides in the patient, where gene expression of at least 60% of the genes correlates with that of the gene signature. The invention further comprises: a treatment of rheumatoid arthritis; identification of genes for targeting in the treatment of rheumatoid arthritis in a mammal other than a mouse; diagnosis of rheumatoid arthritis in a mammal; an array or gene chip, specific for rheumatoid arthritis; diagnosis or analysis of autoimmune disease or rheumatoid arthritis; screening the efficacy of a candidate drug in vitro for the treatment of collagen-induced arthritis; and reducing the symptoms associated with collagen-induced arthritis. The compositions of the invention have the following activities: immunosuppressive, antirheumatic, antiarthritic, osteopathic, antiinflammatory, dermatological, and immunomodulatory. The methods and compositions of the present invention are useful for diagnosing and treating autoimmune disease or arthritides, such as rheumatoid arthritis, lupus, ankylosing spondylitis, fibrositis, fibromyalgia, osteoarthritis, gout, juvenile rheumatoid arthritis, and an immune disease caused by an infectious agent. This sequence represents a protein sequence relating to the genes used in the analysis and treatment of autoimmune diseases or arthritides. Note: This sequence is not shown in the specification. It has been supplied in an electronic format from WIP.

Sequence 331 AA;

Query Match 99.7%; Score 1800; DB 7; Length 331;
Best Local Similarity 99.7%; Pred. No. 1.2e-170;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MKRLVCLVLCSSAVALHDKPTLDHMHLMKTYGQYKKEKEAVRRLIWEKNLKFVM 60
DB 1 MKRLVCLVLCSSAVALHDKPTLDHMHLMKTYGQYKKEKEAVRRLIWEKNLKFVM 60
QY 61 LHNLEHSGMHSYDLGNMHLGDMTSEEVMSLSLRVPSQWRNITYKSNPNRLLPDSVD 120
DB 61 LHNLEHSGMHSYDLGNMHLGDMTSEEVMSLSLRVPSQWRNITYKSNPNRLLPDSVD 120
QY 121 WRKGCVTEVYKQSCGCAWAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTERYKNGKC 180
DB 121 WRKGCVTEVYKQSCGCAWAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTERYKNGKC 180
QY 121 WRKGCVTEVYKQSCGCAWAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTERYKNGKC 180
DB 121 WRKGCVTEVYKQSCGCAWAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTERYKNGKC 180
QY 181 NGFMFTTAFQYIINDKIGIDSDASYPKAMDCKQYDSKRYRAATCSKYTELPYGRDVDLKE 240
DB 181 NGFMFTTAFQYIINDKIGIDSDASYPKAMDCKQYDSKRYRAATCSKYTELPYGRDVDLKE 240
QY 241 AVANKGPVSVGDARHPSFPLYRSGVYEPSCQNVNMGVLVVGDLNGKEYMLVKNWSW 300
DB 241 AVANKGPVSVGDARHPSFPLYRSGVYEPSCQNVNMGVLVVGDLNGKEYMLVKNWSW 300
QY 301 GHNFGEBGYIRMARKNHCGIASFPSPYPEI 331
DB 301 GHNFGEBGYIRMARKNHCGIASFPSPYPEI 331

```

RESULT 4
ADY14944
ID ADY14944 standard; protein; 331 AA.

AC ADY14944;

DT 05-MAY-2005 (first entry)

DE PRO polypeptide SEQ ID NO 750.

XX Antinflammatory; Immune disorder; Dermatological; Immunosuppressive;
XX Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;
XX Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;
XX Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiaesthetic;
XX Antiallergic; ds; gene; diagnosis.

OS Homo sapiens.

XX WO2005016962-A2.

PD 24-FEB-2005.
XX 11-AUG-2004; 2004WO-US026249.
XX 11-AUG-2003; 2003US-0493546P.
XX (GBTH) GENENTECH INC.
PI Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;
XX WPI; 2005-182330/19.
DR
XX New nucleic acid encoding PRO polypeptide, useful for diagnosing and
XX treating an immune related disorder, e.g. systemic lupus erythematosus,
XX rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.
XX Claim 8; SEQ ID NO 750; 158bp; English.
PS
XX The invention relates to an isolated nucleic acid encoding a PRO
XX polypeptide. The polypeptide, agonist or an antagonist, antibody,
XX composition, and method are useful for diagnosing and treating an immune
XX related disorder, e.g. systemic lupus erythematosus, rheumatoid
XX arthritis. The present sequence represents a DNA encoding a PRO
XX polypeptide.

Sequence 331 AA;

Query Match 99.7%; Score 1800; DB 9; Length 331;
Best Local Similarity 99.7%; Pred. No. 1.2e-170;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MKRLVCLVLCSSAVALHDKPTLDHMHLMKTYGQYKKEKEAVRRLIWEKNLKFVM 60
DB 1 MKRLVCLVLCSSAVALHDKPTLDHMHLMKTYGQYKKEKEAVRRLIWEKNLKFVM 60
QY 61 LHNLEHSGMHSYDLGNMHLGDMTSEEVMSLSLRVPSQWRNITYKSNPNRLLPDSVD 120
DB 61 LHNLEHSGMHSYDLGNMHLGDMTSEEVMSLSLRVPSQWRNITYKSNPNRLLPDSVD 120
QY 121 WRKGCVTEVYKQSCGCAWAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTERYKNGKC 180
DB 121 WRKGCVTEVYKQSCGCAWAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTERYKNGKC 180
QY 121 WRKGCVTEVYKQSCGCAWAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTERYKNGKC 180
DB 121 WRKGCVTEVYKQSCGCAWAFSAVGALEAOLKLTGTLVLSAQNLDVDCSTERYKNGKC 180
QY 181 NGFMFTTAFQYIINDKIGIDSDASYPKAMDCKQYDSKRYRAATCSKYTELPYGRDVDLKE 240
DB 181 NGFMFTTAFQYIINDKIGIDSDASYPKAMDCKQYDSKRYRAATCSKYTELPYGRDVDLKE 240
QY 241 AVANKGPVSVGDARHPSFPLYRSGVYEPSCQNVNMGVLVVGDLNGKEYMLVKNWSW 300
DB 241 AVANKGPVSVGDARHPSFPLYRSGVYEPSCQNVNMGVLVVGDLNGKEYMLVKNWSW 300
QY 301 GHNFGEBGYIRMARKNHCGIASFPSPYPEI 331
DB 301 GHNFGEBGYIRMARKNHCGIASFPSPYPEI 331

```

RESULT 5
ADY14946

ID ADY14946 standard; protein; 331 AA.

AC ADY14946;

DT 05-MAY-2005 (first entry)

DE PRO polypeptide SEQ ID NO 752.

XX Antinflammatory; Immune disorder; Dermatological; Immunosuppressive;
XX Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;
XX Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;
XX Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiaesthetic;
XX Antiallergic; ds; gene; diagnosis.

OS Homo sapiens.

XX WO2005016962-A2.
XX
XX 24-FEB-2005.
XX
XX 11-AUG-2004; 2004WO-US026249.
XX
XX 11-AUG-2003; 2003US-0493546P.
XX
XX (GETH) GENENTECH INC.
XX
XX Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;
XX WPI; 2005-182330/19.
XX
XX New nucleic acid encoding PRO polypeptide, useful for diagnosing and
XX treating an immune related disorder, e.g. systemic lupus erythematosus,
XX rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.
XX
XX Claim 8; SEQ ID NO 752; 158bp; English.

XX Sequence 331 AA;

Query Match 99.7%; Score 1800; DB 9; Length 331;
Best Local Similarity 99.7%; Pred. No. 1.2e-170;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKRLVCTLVNCSAVALHDKPTLDHMHLMKTKYKQYKKNBEAVRLLIMEKXKTFVM 60
DB 1 MKRLVCTLVNCSAVALHDKPTLDHMHLMKTKYKQYKKNBEAVRLLIMEKXKTFVM 60
QY 61 LHNLEHSGMHSYDLGMNHLGDMTSEEVMSLSLRVPSOMQRNTTYSNPRILIPDSVD 120
DB 61 LHNLEHSGMHSYDLGMNHLGDMTSEEVMSLSLRVPSOMQRNTTYSNPRILIPDSVD 120
QY 121 WREKCVTEVVKYQSGCAWAFSAVGALEAOLKLTGKLVSLSAQNLVDCSTEKYGNKGC 180
DB 121 WREKCVTEVVKYQSGCAWAFSAVGALEAOLKLTGKLVSLSAQNLVDCSTEKYGNKGC 180
QY 181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKTYRAATGSKYTELPYGRDVLKE 240
DB 181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKTYRAATGSKYTELPYGRDVLKE 240
QY 241 AVANKGPVSVGVDAHPSFFLYRSGVYEPSCQOVNHNGLVVGVDLNGKEXYMLVKNKM 300
DB 241 AVANKGPVSVGVDAHPSFFLYRSGVYEPSCQOVNHNGLVVGVDLNGKEXYMLVKNKM 300
QY 301 GHNFGEBGYIRMARKNHCGIASPSPYPEI 331
DB 301 GHNFGEBGYIRMARKNHCGIASPSPYPEI 331

RESULT 6
ID AEC01587 standard; protein; 331 AA.

XX AEC01587;

XX 20-OCT-2005 (first entry)

XX Human cathepsin S (CTSS) protein, SEQ ID NO: 20.

XX Alzheimer's disease; neuroprotective; neurotropic; degeneration;
XX neurological disease; neurodegenerative disease; diagnosis; gene therapy;
XX prognosis; cathepsin S.

OS Homo sapiens.

XX US2005176030-A1.

XX 11-AUG-2005.

XX 25-OCT-2004; 2004US-00973858.

XX 28-OCT-2003; 2003US-0515562P.

XX (GANL/) GAN L.

XX (GONZ/) GONZALEZ-ZUJUETA M.

XX (YESS/) YE S.

XX (URFE/) URFER R.

XX (NIKO/) NIKOLICH K.

XX Gan L, Gonzalez-Zulueta M, Ye S, Urfer R, Nikolich K;

XX WPI; 2005-581186/59.

XX N-PSDB; AEC01586.

XX REFSQ; NP_004070.

XX Claim 1; SEQ ID NO 20; 148bp; English.

XX The present invention relates to a method of detecting a
XX neurodegenerative disorder (preferably Alzheimer's disease). The method
XX involves detecting the presence of differential expression of a gene
XX encoding a polypeptide having a linear peptide sequence in biological
XX sample. The invention is also useful in gene therapy. The present
XX sequence is the human cathepsin S (CTSS) protein. This protein encoding
XX gene is one of the Alzheimer's disease associated gene.

XX Sequence 331 AA;

Query Match 99.7%; Score 1800; DB 9; Length 331;
Best Local Similarity 99.7%; Pred. No. 1.2e-170;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKRLVCTLVNCSAVALHDKPTLDHMHLMKTKYKQYKKNBEAVRLLIMEKXKTFVM 60
DB 1 MKRLVCTLVNCSAVALHDKPTLDHMHLMKTKYKQYKKNBEAVRLLIMEKXKTFVM 60
QY 61 LHNLEHSGMHSYDLGMNHLGDMTSEEVMSLSLRVPSOMQRNTTYSNPRILIPDSVD 120
DB 61 LHNLEHSGMHSYDLGMNHLGDMTSEEVMSLSLRVPSOMQRNTTYSNPRILIPDSVD 120
QY 121 WREKCVTEVVKYQSGCAWAFSAVGALEAOLKLTGKLVSLSAQNLVDCSTEKYGNKGC 180
DB 121 WREKCVTEVVKYQSGCAWAFSAVGALEAOLKLTGKLVSLSAQNLVDCSTEKYGNKGC 180
QY 181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKTYRAATGSKYTELPYGRDVLKE 240
DB 181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKTYRAATGSKYTELPYGRDVLKE 240
QY 241 AVANKGPVSVGVDAHPSFFLYRSGVYEPSCQOVNHNGLVVGVDLNGKEXYMLVKNKM 300
DB 241 AVANKGPVSVGVDAHPSFFLYRSGVYEPSCQOVNHNGLVVGVDLNGKEXYMLVKNKM 300
QY 301 GHNFGEBGYIRMARKNHCGIASPSPYPEI 331
DB 301 GHNFGEBGYIRMARKNHCGIASPSPYPEI 331

RESULT 7
ID ADD45416 standard; protein; 331 AA.

XX ADD45416;

DT 29-JAN-2004 (first entry)
 XX Human Protein P25774, SEQ ID NO 10849.
 DE
 XX Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SN1; Chung.
 XX Homo sapiens.
 OS
 XX WO2003016475-A2.
 PN
 XX 27-FEB-2003.
 PD
 XX 14-AUG-2002; 2002WO-US025765.
 PF
 XX 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX
 XX WPI; 2003-268312/26.
 DR GENBANK; P25774.
 DR
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antihoddes. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SN1) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 331 AA;
 Query Match 99.5%; Score 1797; DB 7; Length 331;
 Best Local Similarity 99.4%; Pred. No. 2,4e-170; Indels 0; Gaps 0;
 Matches 329; Conservative 1; Mismatches 1;
 QY 1 MKRLVCLLVGSSAAYALHKOPTLHHHMLMKQYKQYKKNRBAVRRLIWEKXUKFVM 60
 DB 1 MKRLVCLLVGSSAAYALHKOPTLHHHMLMKQYKQYKKNRBAVRRLIWEKXUKFVM 60
 QY 61 LHNLEHSGMHSYDLGNMHLGDMTSEEVWSLMSLRVPSQWRNITTKYKSNPRLIPDSVD 120

DB 61 LHNLEHSGMHSYDLGNMHLGDMTSEEVWSLMSLRVPSQWRNITTKYKSNPRLIPDSVD 120
 QY WREKGVTEVYKQSCGCMAPFAVGALEAQLKLTGTVSIAQNLVDCSTKXGNKGC 180
 DB WREKGVTEVYKQSCGCMAPFAVGALEAQLKLTGTVSIAQNLVDCSTKXGNKGC 180
 QY 181 NGFMETAFQYIINDKNGIDSDASYPKAMDLCQYDSKYRAATCSKYTELPYGRDVLKE 240
 DB 181 NGFMETAFQYIINDKNGIDSDASYPKAMDLCQYDSKYRAATCSKYTELPYGRDVLKE 240
 QY 241 AVANKGPVSVGDARHPSPFLYRSGVYBESCTQNVNMGVTVVYGGDLNGEYHLYKNSW 300
 DB 241 AVANKGPVSVGDARHPSPFLYRSGVYBESCTQNVNMGVTVVYGGDLNGEYHLYKNSW 300
 QY 301 GNFGEEGYIMARKNGHCGIASFPSYPEI 331
 DB 301 GNFGEEGYIMARKNGHCGIASFPSYPEI 331
 RESULT 8
 ADE56902
 ID ADE56902 standard; protein; 331 AA.
 XX
 AC ADE56902;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human Protein P25774, SEQ ID NO 2757.
 XX
 KW Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SN1; Chung.
 XX
 OS Homo sapiens.
 XX
 XX WO2003016475-A2.
 PN
 XX 27-FEB-2003.
 PD
 XX 14-AUG-2002; 2002WO-US025765.
 PF
 XX 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX
 XX WPI; 2003-268312/26.
 DR GENBANK; P25774.
 DR
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the

CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 331 AA;

Query Match 99.5%; Score 1797; DB 7; Length 331;

Best Local Similarity 99.4%; Pred. No. 2.4e-170; Mismatches 1; Indels 0; Gaps 0;

Matches 329; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVCCSAVAQJHKDPTLDHMHLMKTYGKQYKEKNEAVRLIWEKNLKFVM 60
 DB 1 MKRLVCVLLVCCSAVAQJHKDPTLDHMHLMKTYGKQYKEKNEAVRLIWEKNLKFVM 60
 QY 61 LHNLEHSMGMSHYDGLGNHLDGMTSEEVMSLSLRVPSQOMORNTTYSNPRILPDSVD 120
 DB 61 LHNLEHSMGMSHYDGLGNHLDGMTSEEVMSLSLRVPSQOMORNTTYSNPRILPDSVD 120
 QY 121 WREKCVTEVKYQSGCGACMAFSAVGALEAOLKLTGKLVLSAONLVDCSTEKYGNKGC 180
 DB 121 WREKCVTEVKYQSGCGACMAFSAVGALEAOLKLTGKLVLSAONLVDCSTEKYGNKGC 180
 QY 181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKRYRATCSKYTELPYGRBDVLKE 240
 DB 181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKRYRATCSKYTELPYGRBDVLKE 240
 QY 241 AVANGPVSVGVDAHRPFILYRSGVYIEPCTQNVNNGVLVVGDLNGKEXYMLVKNSW 300
 DB 241 AVANGPVSVGVDAHRPFILYRSGVYIEPCTQNVNNGVLVVGDLNGKEXYMLVKNSW 300
 QY 301 GHNFGEGYIMARKNGHCGIASFPSYPEI 331
 DB 301 GHNFGEGYIMARKNGHCGIASFPSYPEI 331

RESULT 9

ADJ71695 ID ADJ71695 standard; protein; 331 AA.

AC ADJ71695;

DT 06-MAY-2004 (first entry)

DE Human NOV9a protein SEQ ID NO:96.

KW human; cytosolic; antidiabetic; anorectic; CNS; cardiovascular;
 KW antiinflammatory; gene therapy; antisense therapy; cancer; diabetes;
 KW obesity; endocrine disorder; inflammatory disorder.

OS Homo sapiens.

PN WO2004015076-A2.

XX 19-FEB-2004.

XX 07-AUG-2003; 2003WO-US024788.

XX 07-AUG-2002; 2002US-0401597P.

XX 09-AUG-2002; 2002US-0402248P.

XX 12-AUG-2002; 2002US-0402815P.

XX 13-AUG-2002; 2002US-0403485P.

XX 14-AUG-2002; 2002US-0403574P.

XX 15-AUG-2002; 2002US-0403732P.

XX 20-AUG-2002; 2002US-0404829P.

PR 27-AUG-2002; 2002US-0406392P.

PR 06-AUG-2003; 2003US-00406392.

XX (CURA-) CURAGEN CORP.

PI Anderson DW, Berghe C, Catterton E, Edinger SR, Gorman L, Guo X;

PI Herrmann JL, Kexuda R, Li L, Rieger DK, Zhong M;

DR WPI; 2004-180659/17.

DR N-PSDB; ADJ71694.

PT Novel polypeptides (NOVX) and nucleic acid molecules useful for treating,

PT preventing and diagnosing pathological conditions with NOVX-associated

PT disorders, such as cancer, obesity, diabetes and inflammatory diseases.

PS Claim 2; SEQ ID NO 96; 267pp; English.

CC The invention relates to a novel isolated NOVX polypeptide. A polypeptide
 CC of the invention has cytosolic, antidiabetic, anorectic, CNS-gen,
 CC cardiovascular-gen, and antiinflammatory activity. A polynucleotide
 CC encoding a polypeptide of the invention may have a use in gene therapy,
 CC and antisense therapy. The methods and compositions of the present
 CC invention are useful for the diagnosis and treatment of disorders
 CC associated with aberrant expression or activity of the NOVX polypeptide,
 CC such as cancer, diabetes, obesity, and endocrine, CNS, cardiovascular and
 CC inflammatory disorders. They can also be used in various detection and
 CC screening assays, chromosome mapping, tissue typing and predictive
 CC medicine. The present sequence represents a NOVX polypeptide of the
 CC invention.

CC Sequence 331 AA;

Query Match 99.5%; Score 1797; DB 8; Length 331;

Best Local Similarity 99.4%; Pred. No. 2.4e-170; Mismatches 1; Indels 0; Gaps 0;

Matches 329; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKRLVCVLLVCCSAVAQJHKDPTLDHMHLMKTYGKQYKEKNEAVRLIWEKNLKFVM 60
 DB 1 MKRLVCVLLVCCSAVAQJHKDPTLDHMHLMKTYGKQYKEKNEAVRLIWEKNLKFVM 60
 QY 61 LHNLEHSMGMSHYDGLGNHLDGMTSEEVMSLSLRVPSQOMORNTTYSNPRILPDSVD 120
 DB 61 LHNLEHSMGMSHYDGLGNHLDGMTSEEVMSLSLRVPSQOMORNTTYSNPRILPDSVD 120
 QY 121 WREKCVTEVKYQSGCGACMAFSAVGALEAOLKLTGKLVLSAONLVDCSTEKYGNKGC 180
 DB 121 WREKCVTEVKYQSGCGACMAFSAVGALEAOLKLTGKLVLSAONLVDCSTEKYGNKGC 180
 QY 181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKRYRATCSKYTELPYGRBDVLKE 240
 DB 181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKRYRATCSKYTELPYGRBDVLKE 240
 QY 241 AVANGPVSVGVDAHRPFILYRSGVYIEPCTQNVNNGVLVVGDLNGKEXYMLVKNSW 300
 DB 241 AVANGPVSVGVDAHRPFILYRSGVYIEPCTQNVNNGVLVVGDLNGKEXYMLVKNSW 300
 QY 301 GHNFGEGYIMARKNGHCGIASFPSYPEI 331
 DB 301 GHNFGEGYIMARKNGHCGIASFPSYPEI 331

RESULT 10

ADD35927 ID ADD35927 standard; peptide; 331 AA.

XX ADD35927;

XX 15-JAN-2004 (first entry)

XX Human cathepsin S wild-type.

XX Crystal; cathepsin S; cats inhibitor; human.

Query Match	1794	DB 7	length 331
Best Local Similarity	99.4%	Pred No. 4.8e-170	
Matches 329	Conservative 0	Mismatches 2	Indels 0
			Gaps 0

RESULT 11

```
DT      05-MAY-2005  (first entry)
XX
DE      PRO polypeptide SEQ ID NO 5590
..
```

XX

Query Match	99.2%	Score 1791;	DB 9;	Length 331;
Best Local Similarity	99.1%;	Pred. No. 9.5e-170;		
Matches 328;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0

RESULT 12
ADD3E977

AC	ADD35933;
XX	
DT	15-JAN-2004 (first entry)
...	

OS Homo sapiens.

XX US2003143714-A1.
 PN 31-JUL-2003.
 XX 18-OCT-2002; 2002US-00273577.
 XX 18-OCT-2002; 2002US-00273577.
 XX 19-OCT-2001; 2001US-0330191P.
 XX (MEDI-) MEDIVIR UK LTD.
 XX Lamers MBAC, Williams DH, Turkenburg JP, Hubbard RE;
 XX WPI, 2003-829792/77.
 XX
 XX Crystalline cathepsin S polypeptide free of irreversible inhibitor useful
 PT for identifying and producing potential cats inhibitor.
 XX
 XX Claim 2; SEQ ID NO 7; 59pp; English.
 XX
 XX The invention relates to a crystalline cathepsin S polypeptide. The
 CC crystalline cats polypeptide is useful for identifying a potential cats
 CC inhibitor molecule. The present sequence represents the amino acid
 CC sequence of a human cathepsin S mutant.
 XX
 XX Sequence 331 AA;
 SQ

Query Match 99.1%; Score 1790; DB 7; Length 331;
 Best Local Similarity 99.4%; Pred. No. 1.2e-169;
 Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCLLVGSSAVALQHKDPTLDHWHLMKTYGKQYKNEBAVRLLIWEKXLFVM 60
 DB 1 MKRLVCLLVGSSAVALQHKDPTLDHWHLMKTYGKQYKNEBAVRLLIWEKXLFVM 60
 QY 61 LHNLESHGMSYDGMNHLGDMTSEEVMSLSLRVPSQORNTTYSNPRILPDSVD 120
 DB 61 LHNLESHGMSYDGMNHLGDMTSEEVMSLSLRVPSQORNTTYSNPRILPDSVD 120
 QY 121 WREKCVTEVRYQSGCGAFSAVGALEAOLKLTGKLVLSAQNLYDCSTEKYGNKGC 180
 DB 121 WREKCVTEVRYQSGCGAFSAVGALEAOLKLTGKLVLSAQNLYDCSTEKYGNKGC 180
 QY 181 NGFMTTAFQYIINDKGISDASYPKAMDLCQYDSKYRAATCSKYTELPYGRDVLKE 240
 DB 181 NGFMTTAFQYIINDKGISDASYPKAMDLCQYDSKYRAATCSKYTELPYGRDVLKE 240
 QY 241 AVANKGPVSVGDARHPFFLYRSGVYEPSCQVNVNNGVLVVGDLNGKEXYMLVKNSW 300
 DB 241 AVANKGPVSVGDARHPFFLYRSGVYEPSCQVNVNNGVLVVGDLNGKEXYMLVKNSW 300
 QY 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331
 DB 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331

RESULT 13
 ADD35934
 ID ADD35934 standard; peptide; 331 AA.
 XX
 XX ADD35934;
 AC
 XX
 XX 15-JAN-2004 (first entry)
 DT
 XX
 XX Human cathepsin S mutant #3.
 DE
 XX
 XX Crystalline cathepsin S; cats inhibitor; human; mutant; mutein.
 KW
 XX
 XX Synthetic.
 OS
 XX Homo sapiens.
 XX US2003143714-A1.
 XX

PD 31-JUL-2003.
 XX 18-OCT-2002; 2002US-00273577.
 XX 18-OCT-2002; 2002US-00273577.
 XX 19-OCT-2001; 2001US-0330191P.
 XX (MEDI-) MEDIVIR UK LTD.
 XX Lamers MBAC, Williams DH, Turkenburg JP, Hubbard RE;
 XX WPI, 2003-829792/77.
 XX
 XX Crystalline cathepsin S polypeptide free of irreversible inhibitor useful
 PT for identifying and producing potential cats inhibitor.
 XX
 XX Claim 2; SEQ ID NO 8; 59pp; English.
 XX
 XX The invention relates to a crystalline cathepsin S polypeptide. The
 CC crystalline cats polypeptide is useful for identifying a potential cats
 CC inhibitor molecule. The present sequence represents the amino acid
 CC sequence of a human cathepsin S mutant.
 XX
 XX Sequence 331 AA;
 SQ

Query Match 98.8%; Score 1784; DB 7; Length 331;
 Best Local Similarity 99.1%; Pred. No. 4.8e-169;
 Matches 328; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKRLVCLLVGSSAVALQHKDPTLDHWHLMKTYGKQYKNEBAVRLLIWEKXLFVM 60
 DB 1 MKRLVCLLVGSSAVALQHKDPTLDHWHLMKTYGKQYKNEBAVRLLIWEKXLFVM 60
 QY 61 LHNLESHGMSYDGMNHLGDMTSEEVMSLSLRVPSQORNTTYSNPRILPDSVD 120
 DB 61 LHNLESHGMSYDGMNHLGDMTSEEVMSLSLRVPSQORNTTYSNPRILPDSVD 120
 QY 121 WREKCVTEVRYQSGCGAFSAVGALEAOLKLTGKLVLSAQNLYDCSTEKYGNKGC 180
 DB 121 WREKCVTEVRYQSGCGAFSAVGALEAOLKLTGKLVLSAQNLYDCSTEKYGNKGC 180
 QY 181 NGFMTTAFQYIINDKGISDASYPKAMDLCQYDSKYRAATCSKYTELPYGRDVLKE 240
 DB 181 NGFMTTAFQYIINDKGISDASYPKAMDLCQYDSKYRAATCSKYTELPYGRDVLKE 240
 QY 241 AVANKGPVSVGDARHPFFLYRSGVYEPSCQVNVNNGVLVVGDLNGKEXYMLVKNSW 300
 DB 241 AVANKGPVSVGDARHPFFLYRSGVYEPSCQVNVNNGVLVVGDLNGKEXYMLVKNSW 300
 QY 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331
 DB 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331

RESULT 14
 AA030451
 ID AA030451 standard; protein; 330 AA.
 XX
 XX AA030451;
 AC
 XX
 XX 22-SEP-2003 (first entry)
 DT
 XX
 XX Monkey cathepsin S protein.
 DE
 XX
 XX Monkey; cathepsin S; immune disease; lupus; rheumatoid arthritis; asthma;
 KW gene therapy.
 XX
 XX Macaca sp.
 OS
 XX
 XX WO2003040396-A2.
 PN
 XX
 XX 15-MAY-2003.
 PD
 XX
 XX 25-OCT-2002; 2002WO-US034383.
 PF

XX 08-NOV-2001; 2001US-00010577.
 XX (ORTH) ORTHO-MCNEIL PHARM INC.
 XX Thurmond R, Karlsson L, Baker S;
 XX WPI; 2003-430675/40.
 DR N-PSDB; AAL61281.
 XX
 PT New nucleic acid molecule encoding a monkey cathepsin S protein, useful
 PT for identifying modulators of cathepsin S for treating chronic immune
 PT diseases such as lupus, rheumatoid arthritis, or asthma.
 PS
 PS Claim 6; fig 2; 57pp; English.
 XX
 CC The invention relates to monkey cathepsin S proteins and nucleic acid
 CC molecules encoding such proteins. Polypeptides of the invention are
 CC useful for identifying cathepsin S modulators which are useful as
 CC therapeutic agents for treating conditions mediated by cathepsin S e.g.
 CC chronic immune diseases such as lupus, rheumatoid arthritis or asthma.
 CC The invention is also useful in gene therapy. The present sequence is
 CC monkey cathepsin S protein
 XX
 SQ Sequence 330 AA;
 Query Match 94.2%; Score 1700.5; DB 6; Length 330;
 Best Local Similarity 93.7%; Pred. No. 1e-160; Indels 1; Gaps 1;
 Matches 310; Conservative 11; Mismatches 9;
 QY 1 MKRLVCTLVCSAVALHDKPTLDHNMWLMKTYGKQYKKNBEAVRLIWEKNLKFVM 60
 DB 1 MKQVLCVLPVCSAVALTQHLKOPTLDHNMWLMKTYGKQYKKNBEAVRLIWEKNLKFVM 60
 QY 61 LHNLEHSGMHSYDLGNMHLGDMTSEEVVMSLSLRVPSQWQRNTTYSNPNRIIPDSVD 120
 DB 61 LHNLEHSGMHSYDLGNMHLGDMTSEEVVMSLSLRVPSQWQRNTTYSNPNRIIPDSVD 120
 QY 121 WREKGCVEVYKQSCGACMAFSAVGALEAQLKLTGTLVSLSAQNLVDCSTEEKYGNKGC 180
 DB 121 WREKGCVEVYKQSCGACMAFSAVGALEAQLKLTGTLVSLSAQNLVDCS-EKYGNKGC 179
 QY 121 WREKGCVEVYKQSCGACMAFSAVGALEAQLKLTGTLVSLSAQNLVDCS-EKYGNKGC 179
 DB 121 WREKGCVEVYKQSCGACMAFSAVGALEAQLKLTGTLVSLSAQNLVDCS-EKYGNKGC 179
 QY 181 NGGFMTTAFQYTIIDKIGIDSDASYPYKAMDLCQYDSKRYATCSKYTELPYGRADVLEKE 240
 DB 180 NGGFMTTAFQYTIIDKIGIDSDASYPYKATDQCKQYDSKRYATCSKYTELPYGRADVLEKE 239
 QY 241 AVANKGPVSVGVDAHPSPFLYRSGVYVEPCTQNVNNGVLVVGVDLNGKEYMLVKNSW 300
 DB 240 AVANKGPVSVGVDAHPSPFLYRSGVYVEPCTQNVNNGVLVVGVDLNGKEYMLVKNSW 299
 QY 301 GHNFGEBGYIMARKNGHCGIASFPSPYPEI 331
 DB 300 GSNFGEQGYIMARKNGHCGIASYSPYPEI 330
 Db
 RESULT 15
 ABM83324
 ID ABM83324 standard; protein; 313 AA.
 AC ABM83324;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Human diagnostic and therapeutic pprotein SEQ ID NO:3573.
 XX
 KM gene therapy; human diagnostic and therapeutic polynucleotide; dthp.
 XX
 OS Homo sapiens.
 XX
 FN WO2004023973-A2.
 XX
 PD 25-MAR-2004.

PF 12-SEP-2003; 2003WO-US028227.
 XX 12-SEP-2002; 2002US-0410259P.
 PR 12-SEP-2002; 2002US-0410260P.
 XX
 PA (INCY-) INCYTE CORP.
 XX
 PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
 PI Hartsborne TA, Suchorolski MT, Altus CM, Plets SJ, Elder LV;
 PI Mooney EM, Deleage AM, Panesar IS, Banville SC, Reddy TP;
 PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gershtin EH;
 PI Petralle CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
 PI Lagace RE, Spito PA, Stewart EA, Wingrove J, Vilt UA, Kirtton ES;
 PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
 PI Patury S, Shi X, Suarez CJ;
 XX
 DR WPI; 2004-329368/30.
 DR N-PSDB; ACN41976.
 XX
 PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
 PT in diagnosing a condition, disease or disorder associated with human
 PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy, or
 PT in gene mapping.
 PS
 PS Claim 27; Page; 190pp; English.
 XX
 CC The invention relates to novel diagnostic and therapeutic polynucleotides
 CC selected from one of the 2722 sequences defined in the specification. A
 CC polynucleotide of the invention may have a use in gene therapy. The human
 CC diagnostic and therapeutic polynucleotides (dthp) or polypeptides may be
 CC used to diagnose a particular condition, disease or disorder associated
 CC with human molecules, e.g. cell proliferative disorders,
 CC autoimmune/inflammatory disorder, developmental disorder, endocrine
 CC disorder, neurological disorders, gastrointestinal disorders, or
 CC infections caused by virus, bacteria, fungi or parasite. The dthp
 CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germ-line
 CC gene therapy. The present sequence represents a dthp protein of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
 XX
 SQ Sequence 313 AA;
 Query Match 92.9%; Score 1678; DB 8; Length 313;
 Best Local Similarity 94.3%; Pred. No. 1.7e-158; Indels 18; Gaps 1;
 Matches 312; Conservative 0; Mismatches 1;
 QY 1 MKRLVCTLVCSAVALHDKPTLDHNMWLMKTYGKQYKKNBEAVRLIWEKNLKFVM 60
 DB 1 MKRLVCTLVCSAVALHDKPTLDHNMWLMKTYGKQYKKNBEAVRLIWEKNLKFVM 60
 QY 61 LHNLEHSGMHSYDLGNMHLGDMTSEEVVMSLSLRVPSQWQRNTTYSNPNRIIPDSVD 120
 DB 61 LHNLEHSGMHSYDLGNMHLGDMTSEEVVMSLSLRVPSQWQRNTTYSNPNRIIPDSVD 120
 QY 121 WREKGCVEVYKQSCGACMAFSAVGALEAQLKLTGTLVSLSAQNLVDCSTEEKYGNKGC 180
 DB 121 WREKGCVEVYKQSCGACMAFSAVGALEAQLKLTGTLVSLSAQNLVDCSTEEKYGNKGC 172
 QY 181 NGGFMTTAFQYTIIDKIGIDSDASYPYKAMDLCQYDSKRYATCSKYTELPYGRADVLEKE 240
 DB 173 -----YIIDKIGIDSDASYPYKAMDQCKQYDSKRYATCSKYTELPYGRADVLEKE 222
 QY 241 AVANKGPVSVGVDAHPSPFLYRSGVYVEPCTQNVNNGVLVVGVDLNGKEYMLVKNSW 300
 DB 223 AVANKGPVSVGVDAHPSPFLYRSGVYVEPCTQNVNNGVLVVGVDLNGKEYMLVKNSW 282
 QY 301 GHNFGEBGYIMARKNGHCGIASFPSPYPEI 331
 DB 283 GHNFGEBGYIMARKNGHCGIASFPSPYPEI 313

Tue Jan 10 12:17:20 2006

us-10-646-470-1.rag

Page 10

Search completed: January 10, 2006, 09:39:57
Job time : 138 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 09:32:19 ; Search time 39 Seconds

(without alignments)
816.609 Million cell updates/sec

Title: US-10-646-470-1

Perfect score: 1806

Sequence: 1 MKRLVCVTLVCSNAVAQLHK.....MARNKNGIASFPSYPEI 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1791	99.2	331	2	cathepsin S (EC 3.1.1.1)
2	1299.5	72.0	330	2	cathepsin S (EC 3.1.1.1)
3	1035	57.3	217	2	cathepsin S (EC 3.1.1.1)
4	952.5	52.7	329	2	cathepsin K (EC 3.1.1.1)
5	951.5	52.7	329	2	cathepsin K (EC 3.1.1.1)
6	927.5	51.4	329	2	cathepsin K (EC 3.1.1.1)
7	865	47.9	334	1	KHMSL
8	859.5	47.6	334	1	KHMSL
9	840.5	46.5	333	1	KHMSL
10	839	46.5	334	1	KHMSL
11	787	43.6	326	2	S19650
12	785.5	43.5	326	2	S53027
13	785	43.5	313	2	S47433
14	780.5	43.2	339	2	A53810
15	776.5	43.0	337	2	T24387
16	768.5	42.6	344	2	UK0366
17	762	42.2	320	2	S19651
18	761.5	42.2	338	2	JC5443
19	746.5	41.3	333	1	IS2525
20	738	40.9	343	2	JC1183
21	733.5	40.6	324	2	S47432
22	728.5	40.3	322	2	S19649
23	724.5	40.1	331	2	JC5441
24	723	40.0	326	2	S43991
25	721.5	40.0	331	2	JC5442
26	696.5	38.6	218	1	KHCHL
27	689	38.2	480	2	T01207
28	683	37.8	326	2	T09259
29	679	37.6	458	1	KHR20A

30	669.5	37.1	218	2	S67481	cathepsin L-like c
31	652	36.1	462	2	JN0719	drought-inducible
32	643.5	35.6	455	2	T12041	lysosomal protein
33	643	35.6	317	2	S44151	cathepsin L (EC 3.1.1.1)
34	639	35.4	427	2	S57776	cathepsin L (EC 3.1.1.1)
35	630.5	34.9	331	2	D86413	cathepsin L (EC 3.1.1.1)
36	629.5	34.9	368	2	S47312	cathepsin L (EC 3.1.1.1)
37	625.5	34.6	374	2	T03941	cathepsin L (EC 3.1.1.1)
38	625	34.6	367	2	T06529	cathepsin L (EC 3.1.1.1)
39	621.5	34.4	454	2	JC4848	cathepsin L (EC 3.1.1.1)
40	619.5	34.3	464	2	S24602	cathepsin L (EC 3.1.1.1)
41	613.5	34.0	471	1	KHR20B	cathepsin L (EC 3.1.1.1)
42	610.5	33.8	466	2	T06416	cathepsin L (EC 3.1.1.1)
43	610	33.8	355	2	T05390	cathepsin L (EC 3.1.1.1)
44	607	33.6	380	1	TACB	cathepsin L (EC 3.1.1.1)
45	606.5	33.6	343	2	D86198	cathepsin L (EC 3.1.1.1)

ALIGNMENTS

RESULT 1

A42482
cathepsin S (EC 3.4.22.27) precursor - human
C/Species: Homo sapiens (man)
C/Date: 04-Mar-1993 #sequence revision 18-Nov-1994 #text_change 09-Jul-2004
C/Accession: A42482; A53625; A42896

R.Shl. G.P.; Munger, J.S.; Meara, J.P.; Rich, D.H.; Chapman, H.A.

J. Biol. Chem. 267, 7258-7262, 1992
A/Title: Molecular cloning and expression of human alveolar macrophage cathepsin S, an c

A/Reference number: A42482; PMID:92218373; PMID:1373132

A/Accession: A42482

A/Molecule type: mRNA

A/Residues: 1-331 <SH2>

A/Cross-references: UNIPROT:P25774; UNIPARC:UPI0000151182; GB:S93414; NID:G248405; PIDN

A/Experimental source: alveolar macrophage

A/Note: sequence extracted from NCBI backbone (NCBIN:93414, NCBI:93443)

R.Shl. G.P.; Webb, A.C.; Foster, K.E.; Knoll, J.H.M.; Lemere, C.A.; Munger, J.S.; Chapma

J. Biol. Chem. 269, 11530-11536, 1994

A/Title: Human cathepsin S: chromosomal localization, gene structure, and tissue distrib

A/Reference number: A53625; PMID:94209337; PMID:8157683

A/Accession: A53625

A/Molecule type: DNA

A/Residues: 1-210, 'H', 212-331 <SH2>

A/Cross-references: UNIPARC:UPI0000175CF0; GB:U07374

R.Wiederanders, B.; Bromme, D.; Kirsche, H.; von Figura, K.; Schmidt, B.; Peters, C.

J. Biol. Chem. 267, 13708-13713, 1992

A/Title: Phylogenetic conservation of cysteine proteinases. Cloning and expression of a

A/Reference number: A42896; PMID:92317106; PMID:1377692

A/Accession: A42896

A/Status: not compared with conceptual translation

A/Molecule type: mRNA, protein

A/Residues: 1-91, 'W', 93-160, 'S', 162-331 <MT>

A/Cross-references: UNIPARC:UPI000013DFE1; GB:S93127; GB:M90696; NID:G250802; PIDN:AAA09

A/Experimental source: testis

A/Note: sequence extracted from NCBI backbone (NCBI:107806)

C/Genetics:

A/Gene: GDB:CTSS

A/Cross-references: GDB:132414; OMIM:116845

A/Map position: 1q21-1q21

C/Superfamily: papain

C/Keywords: cysteine proteinase; hydrolase; lysosome

F.1-16/Domain: signal sequence #statut predicted <SIG>

F.17-114/Domain: propeptide #statut predicted <PRO>

F.115-331/Product: cathepsin S #statut predicted <MT>

F.1139,278,298/Active site: Cys, His, Asn #statut predicted

Query Match 99.2% Score 1791; DB 2; Length 331;

Best Local Similarity 99.1% Pred. No. 6.2e-142;

Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 MKRLVCVTLVCSNAVAQLHKPTLDHMHMKKTYGKYKNEAVALRLIWRNLFKM 60

Db 1 MKRLVCLLVCSAVALOHKDPILLDHMHMLWKTYGQYKKNBEAVRLLWEKNLKEFVM 60
QY 61 LHNLESHMGMSYDGLMNLGDMTSEBVMWSLSLVPSPOMORNTTYSNPRILLPDSVD 120
Db 61 LHNLESHMGMSYDGLMNLGDMTSEBVMWSLSLVPSPOMORNTTYSNPRILLPDSVD 120
QY 121 WREKCVTEVTKYQSGCGACWAFSAVGALEAOLKLTGKLVLSAONLVDCSTEKYGNKC 180
Db 121 WREKCVTEVTKYQSGCGACWAFSAVGALEAOLKLTGKLVLSAONLVDCSTEKYGNKC 180
QY 181 NGCFMTTAPQYIINDKGLSDSASYPYKAMDLCQYDSKTRATCSKTYELPYGREDEVKE 240
Db 181 NGCFMTTAPQYIINDKGLSDSASYPYKAMDLCQYDSKTRATCSKTYELPYGREDEVKE 240
QY 241 AVANGPVSVDADAPSPFLYRSGVYEPSCQOVNNGVLVVGDLNGKEYMLVKNSM 300
Db 241 AVANGPVSVDADAPSPFLYRSGVYEPSCQOVNNGVLVVGDLNGKEYMLVKNSM 300
QY 301 GHNFGEGYIRMARKGNHCGIASPSPYPEI 331
Db 301 GHNFGEGYIRMARKGNHCGIASPSPYPEI 331

RESULT 2

A:Accession: A45087
A:Species: Rattus norvegicus (Norway rat)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 09-Jul-2004
C:Accession: A45087
R:Reference: S., Devi, L.
J. Biol. Chem. 267, 26038-26043, 1992
A:Title: Sequence analysis, tissue distribution, and expression of rat cathepsin S.
A:Reference number: A45087; MUID:93100327; PMID:1281481
A:Accession: A45087
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-330 <PEP>
A:Cross-references: UNIPROT:002765; UNIPARC:UPI00001270D0; GB:I03201; NID:g203649; PIDN:
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBI:120879, NCBI:120880)
C:Superfamily: papain
C:Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome
F:137,277,297/Active site: Cys, His, Asn #status predicted

Query Match 72.0%; Score 1299.5; DB 2; Length 330;
Best Local Similarity 76.9%; Pred. No. 7e-101;
Matches 240; Conservative 28; Mismatches 41; Indels 3; Gaps 3;

QY 22 PTLDDHMHMLWKTYGQYKKNBEAVRLLWEKNLKEFVM LHNLESHMGMSYDGLMNLG 81
Db 20 PTLDDHMHMLWKTYGQYKKNBEAVRLLWEKNLKEFVM LHNLESHMGMSYDGLMNLG 79
QY 82 DMTSEBVMWSLSLVPSPOMORNTTYSNPRILLPDSVUMREKGVTEKYGSGCGACMA 141
Db 80 DMTSEBVMWSLSLVPSPOMORNTTYSNPRILLPDSVUMREKGVTEKYGSGCGACMA 139
QY 142 FSAVALLEAQLKLTGKLVLSAONLVDCST-EKYGKNGCNGCFMTTAPQYIINDKGLSDS 200
Db 140 FSAVALLEAQLKLTGKLVLSAONLVDCST-EKYGKNGCNGCFMTTAPQYIINDKGLSDS 198
QY 201 DASYPYKAMDLCQYDSKTRATCSKTYELPYGREDEVKEAVANGPVSVDADAPSPFL 259
Db 199 DASYPYKAMDLCQYDSKTRATCSKTYELPYGREDEVKEAVANGPVSVDADAPSPFL 258
QY 260 FLYRSGVYEPSCQOVNNGVLVVGDLNGKEYMLVKNSMGNHCGEGYIRMARKGNH 319
Db 259 FLYRSGVYEPSCQOVNNGVLVVGDLNGKEYMLVKNSMGNHCGEGYIRMARKGNH 318
QY 320 CGIASPSPYPEI 331
Db 319 CGIASPSPYPEI 330

RESULT 3

S15844

cathepsin S (EC 3.4.22.27) - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004

C:Accession: S15844; S23680; S16972; S23957

R:Reference: S15844; S23680; S16972; S23957

A:Title: The complete amino acid sequence of bovine cathepsin S and a partial sequence o

A:Reference number: S15844; MUID:91257334; PMID:2044774

A:Accession: S15844

A:Molecule type: protein

A:Residues: 1-217 <RT>

A:Cross-references: UNIPROT:P25326; UNIPARC:UPI00001270CE

R:Reference: S15844; S23680; S16972; S23957

A:Title: Primary structure of bovine cathepsin S. Comparison to cathepsins L, H, B and p

A:Reference number: S16972; MUID:91323515; PMID:1864368

A:Accession: S23680

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 22-217 <MT>

A:Cross-references: UNIPARC:UPI000016C2D4; GB:M95211; NID:g162814; PIDN:AAA30435.1; PID:

A:Accession: S16972

A:Molecule type: protein

A:Residues: 1-28; 48-71; 94-104; 107-131; 146-162; 178-217 <M12>

A:Cross-references: UNIPARC:UPI0000175CE7; UNIPARC:UPI0000175CE8; UNIPARC:UPI0000175CE9;

C:Superfamily: papain

C:Keywords: cysteine proteinase; hydrolase

F:112-110,22-66,56-99,158-206/Disulfide bonds: #status predicted

F:125,164,184/Active site: Cys, His, Asn #status predicted

Query Match

57.3%; Score 1035; DB 2; Length 217;
Best Local Similarity 84.8%; Pred. No. 5e-79;
Matches 184; Conservative 20; Mismatches 13; Indels 0; Gaps 0;

QY 115 LPDSVUMREKGVTEKYGSGCGACWAFSAVGALEAOLKLTGKLVLSAONLVDCSTBK 174
Db 1 LPDSVUMREKGVTEKYGSGCGACWAFSAVGALEAOLKLTGKLVLSAONLVDCSTBK 60
QY 175 YGNKNGCNGCFMTTAPQYIINDKGLSDSASYPYKAMDLCQYDSKTRATCSKTYELPYGR 234
Db 61 YGNKNGCNGCFMTTAPQYIINDKGLSDSASYPYKAMDLCQYDSKTRATCSKTYELPYGR 120
QY 235 EDVLKEAVANGPVSVDADAPSPFLYRSGVYEPSCQOVNNGVLVVGDLNGKEYM 294
Db 121 EDVLKEAVANGPVSVDADAPSPFLYRSGVYEPSCQOVNNGVLVVGDLNGKEYM 180
QY 295 LVKNSMGNHCGEGYIRMARKGNHCGIASPSPYPEI 331
Db 181 LVKNSMGNHCGEGYIRMARKGNHCGIASPSPYPEI 217

RESULT 4

A49868

cathepsin K (EC 3.4.22.-) precursor [similarity] - rabbit

N:Alternate names: osteoclast cysteine proteinase OC-2

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 30-Jun-1995 #sequence_revision 30-Jun-1995 #text_change 09-Jul-2004

C:Accession: A49868

R:Reference: A49868

A:Title: Molecular cloning of a possible cysteine proteinase predominantly expressed in

A:Reference number: A49868; MUID:94117413; PMID:8288568

A:Accession: A49868

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-329 <TEZ>

A:Cross-references: UNIPROT:P43236; UNIPARC:UPI00001270B5; GB:D14036; NID:g454186; PIDN:

C:Superfamily: papain

C:Keywords: cysteine proteinase; hydrolase

F:139,276,296/Active site: Cys, His, Asn #status predicted

F:139,276 /Active site: Cys, His, Asn #status predicted

```
Query Match      52.7%; Score 951.5; DB 2; Length 329;  
Beat Local Similarity 54.9%; Pred. No. 8.3e-72;  
Matches 184; Conservative 46; Mismatches 86; Indels 19; Gaps 5;
```

Oy VCULLVSSAAVAQLHKDPTLDHMHMLMKKTGYCKQYKKEKNBEAVRRLIWEKNLKFVMHLNL 64
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db VLLLPVVAFPA--LYPEELIDTHWELMKTTRHQYNKKNDVEISRLLIWEKNLKYSIHNL 62

Oy EHSWMGMSHYDLGMNHLGDMTSEEVMSLSLRYPSONORNTIKSPNNRIIL-----PD 117
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db EASLGVTHTYLAMNHLDGMTSEEVOVKMGTLKPPLSHSR-----SNDTXXIPMEGRAPD 117

Oy SVDMREKCVCETEVRKYQGSCGAWCAPSAGALAEOLKLTKGLVLSSLAQNILDSTERYGN 177
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db SVDYRKKGCVTFPVKNQOGCGSCWAFFSVGALLEGOCLKKTKGLINLASPQNVLCVS---N 174

Oy KGCNGEFTTAQQYIIINDKGIDSDDSYPRKAMDLCQDYDSKRATCACTTELPYGREDV 237
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db DCCGGGYMTNAQQYQOKNKIGDISDEAYPPVGGEESCMNPETGAAACRGYRELPEGENKA 234

Oy LKEAAPANGKPSVGVADARHPSPFLYRSYYEPBPC-TONVNHGVLVGYSDINGEKRWLV 296
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db LKRAVARGPVSVALDALSLTSFQFYSKGYIIDSCSNLNHAIVLANGYGIQKNGKIWI 294

Oy KNWSGHNFEGEEGIRMARNRKNGHCIAFPSYPEEI 331
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db KNSWGEMNWGNKGYILLARKRNACCIALNALASFPM 329

RESULT 6

S74227
cathepsin K (EC 3.4.22.-) precursor - mouse
C.Species: Mus musculus (house mouse)
C.Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C.Accession: S74227
R.Rantakokko, J.; Aro, H.T.; Savontaus, M.; Vuorio, E.
FEBS Lett. 393, 307-313, 1996
A.Title: Mouse cathepsin K: cDNA cloning and predominant expression of the gene in osteo-
A.Reference number: S74227; PMID:96409328; PMID:8814310
A.Accession: S74227
A.Molecule type: mRNA
A.Residues: 1-329 <RAW>
A.Cross-references: UNIPROT:P5097; UNIPARC:UPI000044DBO; EMBL:X94444; NID:g1149524; PI
A.Experimental source: strain cst/bl; tissue type calvaria
C.Superfamily: papain
C.Keywords: cysteine proteinase, glycoprotein, hydrolase
F.I-17/Domains: signal sequence #status predicted <SIG>
F.I-18-114/Domains: propeptide #status predicted <PRO>
F.I-15-329/Product: cathepsin K #status predicted <MAT>
F.I103_213/Binding site: carbohydrate (asn) (covalent) #status predicted
F.I136-177,170-210,269-318/disulfide bonds: #status predicted
F.I139,276,/Active site: Cys, His, Asn #status predicted

Query Match 51.4%; Score 927.5; DB 2; Length 329;
Beat Local Similarity 53.7%; Pred. No. 8.4e-70;
Matches 176; Conservative 49; Mismatches 96; Indels 7; Gaps 4;

Oy VLLVSSAAVAQLHKDPTLDHMHLMKTYTGKQYEKENBEAVRRLIMEKNLKFVWLHNLEH 66
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db VLLLPVVAFPA-LSPHEMDLTOWELMWKTTKQKOYNSKVDELISRRLIWEKNLKOISAHLIEA 64

Oy SMGMSHYDLGMNHLGDMTSEEVMSLSLRVP-SOMQRNTIKSPNNRIILPDSVDMREK 124
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db SLGVTHTYLAMNHLDGMTSEEVOVKMGTLKRPSPRYSDNTLYTPMEGKVPDSIDTRKK 124

Oy GCUTEVRTQSGSCGAWCAPSAVALAEOLKLTKGLVSLSSAQNILVDSTEERKNGKCNGF 184
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db GVTVFPVNOQGGSCWAFSSAGALEGOCLKKTKGLINALASPQNVLCVTENV---GCCGGY 181

Oy MTFAQYITINDKDGSDDSYPRKAMDLCQDYDSKRATCACTTELPYGEDVLKNAVAN 244
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db MTTAFYVVOONGGSDISEDAPFYVGQDES CMNNATAKAKRGYAREIPVGNKKLRAVAR 241

Qy 245 KGPISVGVDMARHPFEFLYRSGVYTPESCTQ--NVNAGLVLYGVDLJNKEXYLVNYSNGHN 303

Db 242 VGRFSTVSIDSLASFOFYRSGVYIDENCDRNNVHAFLVVGQYGRKSKHIIITNSNGES 301

Qy 304 FGEEGYTRMARKNKNGCIASFSPSPETI 331

Db 302 WGNKGVALLANNNKACGITTMAAFPRK 329

RESULT 7

cathepsin L (EC 3.4.22.15) precursor - mouse
N:Alternate names: major excreted protein (MEP); procathepsin L
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
A:Accession: S01177; A34972; A25999; A32333; A45927; S13890; S48734; S64672
R:Protein, B. R.; Gal, S.; Gottesman, M. M.
A:Title: Sequence and expression of the cDNA for MEP (major excreted protein), a transcribed
A:Accession: S01177
A:Reference number: S01177; MUID:88076849; PMID:3689328
A:Molecule type: mRNA
A:Residues: 1-334 <TR01>
A:Cross-references: UNIPROT:P06797; UNIPARC:UPI000003E31; EMBL:X06086; NID:G53046; PIDN:
A:Accession: A34972
A:Molecule type: protein
A:Residues: 18-28 <TR02>
A:Cross-references: UNIPARC:UPI0000172C60
R:Portnoy, D. A.; Erickson, A. H.; Kochan, J.; Ravetch, J. V.; Unkeles, J. C.
J. Biol. Chem. 261, 14697-14703, 1986
A:Title: Cloning and characterization of a mouse cysteine proteinase.
A:Reference number: A25999; MUID:87033683; PMID:3533924
A:Accession: A25999
A:Molecule type: mRNA
A:Residues: 1-334 <P0R>
A:Cross-references: UNIPARC:UPI000003E31; EMBL:J02583; NID:G192681; PIDN:AAA37445.1; PID:
R:Joseph, L. J.; Chang, L. C.; Stamenkovich, D.; Sukhatme, V. P.
J. Clin. Invest. 81, 1621-1629, 1988
A:Title: Complete nucleotide and deduced amino acid sequences of human and murine preprot
A:Reference number: A92768; MUID:88213715; PMID:2835398
A:Accession: A32333
A:Molecule type: mRNA
A:Residues: 1-57, 'I', 59-334 <J0S>
A:Cross-references: UNIPARC:UPI000008330F; GB:M0495; NID:G200500; PIDN:AAA39984.1; PID:
R:Denhardt, D. T.; Hamilton, R. T.; Paretti, C. L. J.; Edwards, D. R.; St-Pierre, R.; Waterm
Cancer Res. 46, 4590-4593, 1986
A:Title: Close relationship of the major excreted protein of transformed murine fibrobl
A:Reference number: A45927; MUID:86271744; PMID:3755373
A:Accession: A45927
A:Molecule type: mRNA
A:Residues: 89-300 <D8N>
A:Cross-references: UNIPARC:UPI000016CEB1; GB:X04392; NID:G53050; PIDN:CAA27980.1; PID:G
R:Stearns, N. A.; Dong, J.; Pan, J. X.; Brenner, D. A.; Sahagian, G. G.
Arch. Biochem. Biophys. 283, 447-457, 1990
A:Title: Comparison of cathepsin L synthesized by normal and transformed cells at the ge
A:Reference number: S13890; MUID:91112761; PMID:2275556
A:Accession: S13890
A:Molecule type: protein
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-334 <STE>
A:Cross-references: UNIPARC:UPI000003E31
R:Lehlich, K.; Komlami, E.
FEBS Lett. 352, 281-284, 1994
A:Title: Multi-step processing of procathepsin L in vitro.
A:Reference number: S48734; MUID:95010724; PMID:7925987
A:Accession: S48734
A:Molecule type: protein
A:Residues: 104-124 <ISH>
A:Cross-references: UNIPARC:UPI0000172C61
R:Juan, D.; Herman, J.; Rodriguez-Lima, F.; Barel, M.; Balbo, M.; Frade, R.
Biochem. J. 312, 961-969, 1995

A:Title: Identification on melanoma cells of p39, a cysteine proteinase that cleaves C3.
A:Reference number: S64672; MUID:96128086; PMID:8554545
A:Accession: S64672
A:Status: preliminary
A:Molecule type: protein
A:Residues: 18-34;273-292;295-313 <JEA>
A:Cross-references: UNIPARC:UPI00000172C62; UNIPARC:UPI00000172C64
C:Function:
A:Description: catalyzes hydrolysis of peptide bonds in proteins
A:Pathway: intracellular protein degradation
A>Note: important role in the lysosomal degradation of proteins
C:Superfamily: papain
C:Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; proteinase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-113/Domain: signal sequence #status predicted <PRO>
F:114-268/Product: cathepsin L heavy chain #status predicted <HCH>
F:291-334/Product: cathepsin L light chain #status predicted <LCH>
F:135-178,169-211,269-322/Disulfide bonds: #status predicted
F:118,276,300/Active site: Cys, His, Asn #status predicted
F:222,268/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

Query Match	47.9%;	Score 865;	DB 1;	Length 334;
Best Local Similarity	-49.1%;	Pred. No. 1.4e-64;		
Matches 168;	Conservative 58;	Mismatches 96;	Indels 20;	Gaps 8;

```

Oy 1 MKRIIVLCVC--SSAVVOLHDPDPLIDHMHLMKTKYGYQYKEKKEEAVRRLIWEKTLFV 59
Db 1 MNLILLAVLCIGTALATPKPDQTFSAEMHQMKSTHRLY--GTNEEVRRAIMEQNMMI 59
Oy 60 MLNLEHSMGMSHYDLGNHLGDMTSEFVMSLMSLRVPSQ-----WQRNTTYSNPRRI 114
Db 60 QLNHGEGYSNQHGSMBENNAFGDMTNEBFRVYVGRQKHKKGRLFEPPLMLK----- 113
Oy 115 LPDSDVDRREKCGVTEVYKQSGSCGACMAFSAAGALAEQHLKTKGLVLSAQNLYDCSTEK 174
Db 114 IPKSDVDRREKCGVTPVNNQGGQGSQMAFSAAGCLBGMFLTKGLKLSLBQNLVDGS--HA 172
Oy 175 YNKGCGNGGEMTTAFQYIITDNKGIDSDASYYYKAMDLCQYDSKTYRAATCSKTYELPYGR 234
Db 173 QGNGCGNGGLMDFQYIKENGSGIDSESYEYKADGSCCKRAEPAVANDGFGVDIIP-Q 231
Oy 235 EDVLCKEAVANGAPVSGVDARHPSEFLRSGVYYEPPSC--TONVNHGLVVGX-----DLN 289
Db 232 EKALMKRAVATGPISTVAMDASHPLQFYSSGIYYEPNSSKNLDHGVLVVGXGYEGTDSN 291
Oy 290 GKQYEWLYNYSWGHNFGEEGYTRMARNKCNHGGIASFPYEPET 331
Db 292 KNKYMVLVNSWGESEMGEGYIKIKAKDRNHCGLATTAASYPAV 333

```

RESULT 8

cathepsin L (BC04.22.15) precursor - rat
 N:Alternate names: cyclic protein-2; major excreted protein (MEP); procathepsin L
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
 C:Accession: S07098; S00155; S02445; A41550; S02446
 R: Ishidoh, K.; Komimami, E.; Suzuki, K.; Katunuma, N.
 FEBS Lett. 259, 71-74, 1989
 A:Title: Gene structure and 5'-upstream sequence of rat cathepsin L.
 A:Reference number: S07098; MUID:50092543; PMID:259113
 A:Accession: S07098
 A:Molecule type: DNA
 A:Residues: 1-334 <ISH1>
 A:Cross-references: UNIPROT:P07154; UNIPARC:UPI0000172C5D; EMBL:X51648; NID:957532
 A:Note: Only part of the nucleotide sequence is given
 R: Ishidoh, K.; Towatari, T.; Imaoh, S.; Kawasaki, H.; Komimami, E.; Katunuma, N.; Suzuki
 FEBS Lett. 223, 69-73, 1987
 A:Title: Molecular cloning and sequencing of cDNA for rat cathepsin L.
 A:Reference number: S00155; MUID:88030047; PMID:3666143
 A:Accession: S00155
 A:Molecule type: mRNA
 A:Residues: 1-30, 'Q', 32-237, 'P', 239-334 <ISH2>

A:Cross-references: UNIPARC:UP10000167AA40; EMBL:Y00697; NID:955887; PION:CA66691.1; PMID:17920261
R:Townsend, T.; Kellum, N.
EBBS Lett. 236, 57-61, 1988
A:Title: Amino acid sequence of rat liver cathepsin L.
A:Reference number: S02445; MUID:88296890; PMID:3402618
A:Accession: S02445
A:Molecule type: protein
A:Residues: 114-288,291-334 <TOW>
A:Cross-references: UNIPARC:UP10000172C5E; UNIPARC:UP10000172C5F
R:Ericsson-Lawence, M.; Zabludoff, S.D.; Wright, W.W.
Mol. Endocrinol. 5, 1789-1798, 1991
A:Title: Cyclic protein-2, a secretory product of rat seroli cells, is the proenzyme for
A:Reference number: A41550; MUID:92168015; PMID:1791830
A:Accession: A41550
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 88-334 <ERI>
A:Cross-references: UNIPARC:UP10000170CE0; GB:S85154; NID:g246147; PION:AA621516.1; PID:
C:Genetics:
A:Introns: 42/3; 83/3; 132/3; 207/3; 262/1; 301/2
C:Complex: heterodimer of disulfide linked chains produced from a single chain precursor
C:Function:
A:Description: catalyzes hydrolysis of peptide bonds in proteins
A:Pathway: intracellular protein degradation
A:Note: important role in the lysosomal degradation of proteins
A:Superfamily: papain
C:Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; protein
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-113/Domain: propeptide #status predicted <PRO>
F:114-288/Product: cathepsin L heavy chain #status experimental <HCH>
F:291-334/Product: cathepsin L light chain #status experimental <LCH>
F:115-178, 169-211, 269-332/disulfide bonds: #status predicted
F:118, 276,300/Active site: Cys, His, Asn #status predicted
F:221/Binding site: carbohydrate (Asn) (covalent) #status experimental

[illegible]

RESULT 9
KHHU.
cathepsin L (BC 3.4.22.15) precursor [validated] - human
N:Alternate names: major excreted protein (MEP); procathepsin L
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 09-Jul-2004
C:Accession: S01002; B32333; S09065; A45043; S00323; B27011; A26069; A32683; D27011; F27011; F27012; F27013; F27014; F27015; F27016; F27017; F27018; F27019; F27020; F27021; F27022; F27023; F27024; F27025; F27026; F27027; F27028; F27029; F27030; F27031; F27032; F27033; F27034; F27035; F27036; F27037; F27038; F27039; F27040; F27041; F27042; F27043; F27044; F27045; F27046; F27047; F27048; F27049; F27050; F27051; F27052; F27053; F27054; F27055; F27056; F27057; F27058; F27059; F27060; F27061; F27062; F27063; F27064; F27065; F27066; F27067; F27068; F27069; F27070; F27071; F27072; F27073; F27074; F27075; F27076; F27077; F27078; F27079; F27080; F27081; F27082; F27083; F27084; F27085; F27086; F27087; F27088; F27089; F27090; F27091; F27092; F27093; F27094; F27095; F27096; F27097; F27098; F27099; F27100; F27101; F27102; F27103; F27104; F27105; F27106; F27107; F27108; F27109; F27110; F27111; F27112; F27113; F27114; F27115; F27116; F27117; F27118; F27119; F27120; F27121; F27122; F27123; F27124; F27125; F27126; F27127; F27128; F27129; F27130; F27131; F27132; F27133; F27134; F27135; F27136; F27137; F27138; F27139; F27140; F27141; F27142; F27143; F27144; F27145; F27146; F27147; F27148; F27149; F27150; F27151; F27152; F27153; F27154; F27155; F27156; F27157; F27158; F27159; F27160; F27161; F27162; F27163; F27164; F27165; F27166; F27167; F27168; F27169; F27170; F27171; F27172; F27173; F27174; F27175; F27176; F27177; F27178; F27179; F27180; F27181; F27182; F27183; F27184; F27185; F27186; F27187; F27188; F27189; F27190; F27191; F27192; F27193; F27194; F27195; F27196; F27197; F27198; F27199; F27200; F27201; F27202; F27203; F27204; F27205; F27206; F27207; F27208; F27209; F27210; F27211; F27212; F27213; F27214; F27215; F27216; F27217; F27218; F27219; F27220; F27221; F27222; F27223; F27224; F27225; F27226; F27227; F27228; F27229; F27230; F27231; F27232; F27233; F27234; F27235; F27236; F27237; F27238; F27239; F27240; F27241; F27242; F27243; F27244; F27245; F27246; F27247; F27248; F27249; F27250; F27251; F27252; F27253; F27254; F27255; F27256; F27257; F27258; F27259; F27260; F27261; F27262; F27263; F27264; F27265; F27266; F27267; F27268; F27269; F27270; F27271; F27272; F27273; F27274; F27275; F27276; F27277; F27278; F27279; F27280; F27281; F27282; F27283; F27284; F27285; F27286; F27287; F27288; F27289; F27290; F27291; F27292; F27293; F27294; F27295; F27296; F27297; F27298; F27299; F27300; F27301; F27302; F27303; F27304; F27305; F27306; F27307; F27308; F27309; F27310; F27311; F27312; F27313; F27314; F27315; F27316; F27317; F27318; F27319; F27320; F27321; F27322; F27323; F27324; F27325; F27326; F27327; F27328; F27329; F27330; F27331; F27332; F27333; F27334; F27335; F27336; F27337; F27338; F27339; F27340; F27341; F27342; F27343; F27344; F27345; F27346; F27347; F27348; F27349; F27350; F27351; F27352; F27353; F27354; F27355; F27356; F27357; F27358; F27359; F27360; F27361; F27362; F27363; F27364; F27365; F27366; F27367; F27368; F27369; F27370; F27371; F27372; F27373; F27374; F27375; F27376; F27377; F27378; F27379; F27380; F27381; F27382; F27383; F27384; F27385; F27386; F27387; F27388; F27389; F27390; F27391; F27392; F27393; F27394; F27395; F27396; F27397; F27398; F27399; F27400; F27401; F27402; F27403; F27404; F27405; F27406; F27407; F27408; F27409; F27410; F27411; F27412; F27413; F27414; F27415; F27416; F27417; F27418; F27419; F27420; F27421; F27422; F27423; F27424; F27425; F27426; F27427; F27428; F27429; F27430; F27431; F27432; F27433; F27434; F27435; F27436; F27437; F27438; F27439; F27440; F27441; F27442; F27443; F27444; F27445; F27446; F27447; F27448; F27449; F27450; F27451; F27452; F27453; F27454; F27455; F27456; F27457; F27458; F27459; F27460; F27461; F27462; F27463; F27464; F27465; F27466; F27467; F27468; F27469; F27470; F27471; F27472; F27473; F27474; F27475; F27476; F27477; F27478; F27479; F27480; F27481; F27482; F27483; F27484; F27485; F27486; F27487; F27488; F27489; F27490; F27491; F27492; F27493; F27494; F27495; F27496; F27497; F27498; F27499; F27500; F27501; F27502; F27503; F27504; F27505; F27506; F27507; F27508; F27509; F27510; F27511; F27512; F27513; F27514; F27515; F27516; F27517; F27518; F27519; F27520; F27521; F27522; F27523; F27524; F27525; F27526; F27527; F27528; F27529; F27530; F27531; F27532; F27533; F27534; F27535; F27536; F27537; F27538; F27539; F27540; F27541; F27542; F27543; F27544; F27545; F27546; F27547; F27548; F27549; F27550; F27551; F27552; F27553; F27554; F27555; F27556; F27557; F27558; F27559; F27560; F27561; F27562; F27563; F27564; F27565; F27566; F27567; F27568; F27569; F27

A:Title: Isolation and sequence of a cDNA for human pro-(cathepsin L).
A:Reference number: S01002; MUID:88339905; PMID:3421948
A:Accession: S01002
A:Molecule type: mRNA
A:Residues: 1-333 <GAL>
A:Cross-references: UNIPARC:UPI000044CAA2; GB:M20496; NID:g809235; PIDN:AAA66974.1; PID:R.Joseph, L.J.; Chang, L.C.; Stamenkovich, D.; Sukhatme, V.P.
J. Clin. Invest. 81, 1621-1629, 1988
A:Title: Complete nucleotide and deduced amino acid sequences of human and murine preproA:Reference number: A92768; MUID:88213715; PMID:2835398
A:Accession: B32333
A:Molecule type: mRNA
A:Residues: 1-333 <JOS>
A:Cross-references: UNIPARC:UPI000004CAA2; GB:M20496; NID:g809235; PIDN:AAA66974.1; PID:R.Joseph, L.; Lapid, S.; Sukhatme, V.
Nucleic Acids Res. 15, 1186, 1987
A:Title: The major rat induced protein in NIH3T3 cells is cathepsin L.
A:Reference number: S09065; MUID:87174843; PMID:3550705
A:Accession: S09065
A:Molecule type: mRNA
A:Residues: 113-154 <O2>
A:Cross-references: UNIPARC:UPI000016A6C; EMBL:X05256; NID:g29718; PIDN:CAA28877.1; PID:J.Chauhan, S.S.; Popescu, N.C.; Ray, D.; Fleischmann, R.; Gottesman, M.M.; Troen, B.R.
J. Biol. Chem. 268, 1039-1045, 1993
A:Title: Cloning, genomic organization, and chromosomal localization of human cathepsinA:Reference number: A45043; MUID:93123212; PMID:8419312
A:Accession: A45043
A:Molecule type: DNA
A:Residues: 40-46;82-86;130-135;205-210;259-264;299-304 <CHA>
A:Cross-references: UNIPARC:UPI0000172C53; UNIPARC:UPI0000172C54; UNIPARC:UPI0000172C55,
A:Note: only exon-intron splice junctions are shown
R.Ritonja, A.; Popovic, T.; Kocnik, M.; Machleidt, W.; Turk, V.
FEBS Lett. 228, 341-345, 1988
A:Title: Amino acid sequences of the human kidney cathepsins H and L.
A:Reference number: S00322; MUID:88137635; PMID:3342889
A:Accession: S00323
A:Molecule type: protein
A:Residues: 114-147, 'P',148-220, 'X',222-267, 'N',269-288;292-333 <RT>
A:Cross-references: UNIPARC:UPI000011D17; UNIPARC:UPI000012C59
R.Machleidt, W.; Ritonja, A.; Popovic, T.; Kocnik, M.; Edzin, J.; Turk, V.; Machleidt, I
in Cysteine proteinases and Their Inhibitors, Turk, V., ed., pp.3-18, Walter de Gruyter,
A:Title: Human cathepsins B, H and L: characterization by amino acid sequences and someA:Reference number: A27011
A:Accession: B27011
A:Molecule type: protein
A:Residues: 'X',115-129, 'W',131-133, 'E',135-141;292-307, 'TD',310-333 <MA2>
A:Cross-references: UNIPARC:UPI0000172C5A; UNIPARC:UPI0000172C5B
R.Mason, R.W.; Walker, J.B.; Northrop, F.D.
Biochem. J. 240, 373-377, 1986
A:Title: The N-terminal amino acid sequences of the heavy and light chains of human cathA:Reference number: A26069; MUID:87127952; PMID:3545185
A:Accession: A26069
A:Molecule type: protein
A:Residues: 114-147, 'P',149-152, 'Y',292-333 <MA>
A:Cross-references: UNIPARC:UPI000011D17; UNIPARC:UPI0000172C5C
R.Smith, S.M.; Gottesman, M.M.
J. Biol. Chem. 264, 20487-20495, 1989
A:Title: Activity and deletion analysis of recombinant human cathepsin L expressed in EsA:Reference number: A32683; MUID:90062183; PMID:2684378
A:Accession: A32683
A:Contents: annotation
A:Genetics:
A:Gene: GDB:CTSL
A:Cross-references: GDB:119824; OMIM:116880
A:Introns: 42/3, 83/3, 132/3, 207/3, 262/1, 301/2
C:Complex: heterodimer of disulfide linked chains produced from a single chain precursor
C:Function:
A:Description: catalyzes hydrolysis of peptide bonds in proteins
A:Pathway: intracellular protein degradation
A:Note: important role in the lysosomal degradation of proteins
C:Superfamily: papain
C:Keywords: cysteine proteinase, glycoprotein, heterodimer, hydrolase, lysosome, protein
/1-17/Domain: signal sequence #status predicted <Sig>

F:18-113/Domain: propeptide #status predicted <PRO>
 F:114-333/Product: cathepsin L #status experimental <MAT>
 F:114-288/Product: cathepsin L heavy chain #status experimental <HCH>
 F:292-333/Product: cathepsin L light chain #status experimental <LCH>
 F:135-178,169-211,269-322/Disulfide bonds: #status predicted
 F:138,276,300/Active site: Cys, His, Asn #status predicted
 F:221/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 46.5%; Score 840.5; DB 1; Length 333;
 Best Local Similarity 48.5%; Pred. No. 1.6e-62;
 Matches 164; Conservative 57; Mismatches 98; Indels 19; Gaps 8;

QY 4 LVCTLVSSSAVAQJHDPPTLDHMHMLKTKYQKQYKKEKNEAARLLIWEKKLKFVMLLN 63
 DB 5 LILAFCGLIASATLTPHSLSEAOVTKYKAMNRLY-GMNEBGMRAVWEKKMKIIEIHN 63
 QY 64 LEHSMGMSYDLGMNHLDMTSEEVMSLSL--RVPSQ---QGNITTKSPNRIILDS 118
 DB 64 QBYRGKHSFTAMANAFGDMTSEBRQVNMNGPQNKPKRGKVFQEPFLFEA-----PRS 117
 QY 119 VDMREKGVTEVYKQSCGACWAFSAVGALEAQLKLTGKLVLSAQNILVDCSTEKYGNK 178
 DB 118 VDMRKGVYTPYKNGQCGSCWAFSATGALBGMFRKTRGLISLEQNILVDCSGPQ-GNE 176
 QY 179 GNGGCFMTTAFQYIITDNKGIDSDASYPYKAMDLCQYDSKTPAATCSKYTELPYREDVL 238
 DB 177 GNGGLMAYAFQYVDNCGIDSEBSYPYEATESCKYKPYKSVANDTGFVDIP-KQEKAL 235
 QY 239 KEAVNKKGPVSGVDARHPSFPLYSRGGVYEPSC-TQNNHGVLVVGYG---DLNGKEY 293
 DB 236 MKAVALTVGGISVAIDAGHESFLPYKGIYFEPDCSEBMDHGVLVVGYGFBESTBSDNKY 295
 QY 294 WLKNSMGNHFGEBEGYIRMARKNKGNHCGIASPSPYIPEI 331
 DB 296 WLKNSMGEEMNGYVMAKORRHCGIASAASPTV 333

RESULT 10

A:Alternate names: major excreted protein (MEP); procathepsin L
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 16-Oct-1996 #sequence_revision 18-Oct-1996 #text_change 09-Jul-2004
 C:Accession: A58195; S59911; S59916
 R:Okamura, N.; Tamura, M.; Uchiyama, Y.; Sugita, Y.; Dacheux, F.; Dacheux, J.
 Biochim. Biophys. Acta 1245, 221-226, 1995
 A:Title: Direct evidence for the elevated synthesis and secretion of procathepsin L in t
 A:Reference number: S59911; MUID:96069841; PMID:7492581
 A:Accession: A58195
 A:Molecule type: mRNA
 A:Residues: 1-334 <OKA>
 A:Cross-references: UNIPROT:028944; UNIPARC:UPI00001270BB; DDBJ:D37917; NID:g710655; PID
 A:Experimental source: epididymis
 A>Note: the authors translated the codon CCG for residue 203 as Thr and CCC for residue
 A:Accession: S59911
 A:Molecule type: protein
 A:Residues: 18-31 <OK2>
 A:Cross-references: UNIPARC:UPI0000175CP1
 C:Function: heterodimer of disulfide linked chains produced from a single chain precursor
 A:Description: catalyzes hydrolysis of peptide bonds in proteins
 A:Pathway: protein degradation
 A>Note: important role in the lysosomal degradation of proteins
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; protein
 F:1-17/Domain: signal sequence #status predicted <SIG>
 F:18-113/Domain: propeptide #status predicted <PRO>
 F:114-333/Product: cathepsin L #status predicted <MAT>
 F:135-178,169-212,270-323/Disulfide bonds: #status predicted
 F:138,277,301/Active site: Cys, His, Asn #status predicted
 F:222,292/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 46.5%; Score 839; DB 2; Length 334;
 Best Local Similarity 48.8%; Pred. No. 2.1e-62;
 Matches 163; Conservative 58; Mismatches 101; Indels 12; Gaps 8;

QY 4 LVCTLVSSSAVAQJHDPPTLDHMHMLKTKYQKQYKKEKNEAARLLIWEKKLKFVMLLN 63
 DB 7 LPAICLGIASAPKL--QONIDADYKKAKATGRILY-GMNEBGMRAVWEKKMKIIEIHN 63
 QY 64 LEHSMGMSYDLGMNHLDMTSEEVMSLSLRVPSQGNRIITTKSPNRIILDSVDMRE 123
 DB 64 QBYSGKHSFTAMANAFGDMTSEBRQVNMNGPQNKPKRGKVFQEPFLFEA-----PRS 122
 QY 124 KCVTEVYKQSCGACWAFSAVGALEAQLKLTGKLVLSAQNILVDCSTEKYKNGCG 183
 DB 123 KGVYAVNKGQCGSCWAFSATGALBGMFRKTRGLISLEQNILVDCSGPQ-GNGCGNG 181
 QY 184 FMTTAFQYIITDNKGIDSDASYPYKAMDLCQYDSKTPAATCSKYTELPYREDVLKEAV 242
 DB 182 LMDNAPQYVDNCGIDSEBSYPYLGRFNSCTYRFGCAANDTGFVDIPQ-REKALMAV 240
 QY 243 ANKGPVSGVDARHPSFPLYSRGGVYEPSC-TQNNHGVLVVGYG---DLNGKEYLVK 297
 DB 241 ATVPGISVAIDAGHSPFYKSGIYDPDCSKLDHGVLVVGYGFBGTDSNSKFWLVK 300
 QY 298 NSMGNHFGEBEGYIRMARKNKGNHCGIASPSPYIPEI 331
 DB 301 NSMGPWGMNGYVMAKQNNHCGISTASAPTV 334

RESULT 11

A:Cysteine proteinase (BC 3.4.22.-) precursor (clone LCP2) - American lobster
 C:Species: Homarus americanus (American lobster)
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004
 C:Accession: S19650; S31655
 R:Raycock, M.V.; Mackay, R.M.; Di Fruecio, M.; Gallant, J.W.
 FEBS Lett. 292, 115-120, 1991
 A:Title: Molecular cloning of three cDNAs that encode cysteine proteinases in the digest
 A:Reference number: S19649; MUID:92070467; PMID:1959590
 A:Accession: S19650
 A:Molecule type: mRNA
 A:Residues: 1-323 <LAA>
 A:Cross-references: UNIPROT:P25782; UNIPARC:UPI0000128CAC; EMBL:X63568; NID:g11052; PIDN
 R:Raycock, M.V.; Mackay, R.M.; Di Fruecio, M.; Gallant, J.W.
 FEBS Lett. 301, 125, 1992
 A:Title: Correction. Molecular cloning of three cDNAs that encode cysteine proteinases 1
 A:Reference number: S31654; MUID:93083613; PMID:1451782
 A:Accession: S31655
 A:Molecule type: mRNA
 A:Residues: 1-323 <LAA>
 A:Cross-references: UNIPARC:UPI0000128CAC; EMBL:X63568; NID:g11052; PIDN:CAA45128.1; PID
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; disulfide bond; hydrolase; zymogen
 F:1-16/Domain: signal sequence #status predicted <SIG>
 F:117-107/Domain: propeptide #status predicted <PRO>
 F:108-323/Product: cysteine proteinase #status predicted <MAT>
 F:128-171,162-204,263-312/Disulfide bonds: #status predicted
 F:131,270,290/Active site: Cys, His, Asn #status predicted

Query Match 43.6%; Score 787; DB 2; Length 323;
 Best Local Similarity 46.6%; Pred. No. 4.5e-58;
 Matches 156; Conservative 55; Mismatches 98; Indels 26; Gaps 5;

QY 5 VCVTLVSSSAVAQJHDPPTLDHMHMLKTKYQKQYKKEKNEAARLLIWEKKLKFVMLLN 64
 DB 3 VAVFLFCVALLA-----AASPMHFHFGKGRQYVDAAEBSYRVRITFEQNKYIFERNK 56
 QY 65 EHSKMSHSYDLGMNHLDMTSEEVMSLSLRVPSQGNRIITTKSPNRIILDSVDMRE 115
 DB 57 KYENGEVTFENLAKMNGDMTSEEFVAVKNG-----NIPRSGAPVSVFPKKTGPQ 107
 QY 116 PDSVDMREKGVTEVYKQSCGACWAFSAVGALEAQLKLTGKLVLSAQNILVDCSTEKY 175

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2006, 09:32:19 ; Search time 45 Seconds
(without alignments)
608.126 Million cell updates/sec

Title: US-10-646-470-1
Perfect score: 1806
Sequence: 1 MKRLVCVLLVCSAVALQHLK.....MARKNKGHGIASFPSYPEI 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/FCIS_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1806	100.0	331	1	US-08-208-007A-8
2	1806	100.0	331	2	US-09-701-685-2
3	1806	100.0	331	2	US-09-290-586A-20
4	1791	99.2	331	2	US-08-860-255A-2
5	1791	99.2	331	2	US-08-915-095A-8
6	1791	99.2	331	2	US-08-798-096-8
7	1791	99.2	331	2	US-08-798-095A-8
8	1791	99.2	331	2	US-09-953-956-8
9	1791	99.2	331	2	US-08-553-125A-8
10	1791	99.2	331	2	US-10-114-464-8
11	1784	98.8	331	1	US-08-330-121B-4
12	1784	98.8	331	1	US-08-536-861-4
13	1784	98.8	331	4	PCT-US95-13820-4
14	1700.5	94.2	330	2	US-10-010-577-2
15	1599	88.5	331	2	US-10-010-580-2
16	952.5	52.7	329	1	US-08-806-959-2
17	952.5	52.7	329	1	US-08-208-007A-7
18	952.5	52.7	329	1	US-08-330-121B-3
19	952.5	52.7	329	2	US-08-915-095A-7
20	952.5	52.7	329	2	US-08-798-096-7
21	952.5	52.7	329	2	US-08-798-095A-7
22	952.5	52.7	329	2	US-09-953-956-7
23	952.5	52.7	329	2	US-08-553-125A-7
24	952.5	52.7	329	2	US-08-536-861-3
25	952.5	52.7	329	2	US-10-114-464-7
26	952.5	52.7	329	4	PCT-US95-13820-3
27	951.5	52.7	329	1	US-08-330-121B-2

28	951.5	52.7	329	1	US-08-852-807-20	Sequence 20, Appl
29	951.5	52.7	329	2	US-08-964-308-4	Sequence 4, Appl
30	951.5	52.7	329	2	US-08-964-313-4	Sequence 4, Appl
31	951.5	52.7	329	2	US-08-860-255A-1	Sequence 1, Appl
32	951.5	52.7	329	2	US-09-069-138-4	Sequence 4, Appl
33	951.5	52.7	329	2	US-08-915-095A-2	Sequence 2, Appl
34	951.5	52.7	329	2	US-08-798-096-2	Sequence 2, Appl
35	951.5	52.7	329	2	US-08-798-095A-2	Sequence 2, Appl
36	951.5	52.7	329	2	US-09-953-956-2	Sequence 2, Appl
37	951.5	52.7	329	2	US-08-553-125A-2	Sequence 2, Appl
38	951.5	52.7	329	2	US-08-536-861-2	Sequence 2, Appl
39	951.5	52.7	329	2	US-10-114-464-2	Sequence 2, Appl
40	951.5	52.7	329	2	US-09-290-586A-19	Sequence 19, Appl
41	951.5	52.7	329	4	PCT-US95-13820-2	Sequence 2, Appl
42	945.5	52.4	329	1	US-08-208-007A-2	Sequence 2, Appl
43	945.5	52.4	329	4	PCT-US94-04781-2	Sequence 2, Appl
44	944.5	52.3	329	2	US-08-684-932A-36	Sequence 36, Appl
45	942.5	52.2	329	2	US-08-964-308-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-208-007A-8
Sequence 8, Application US/08208007A
Patent No. 5501969
GENERAL INFORMATION
APPLICANT: HASTINGS, ET AL.
TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELIA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/208,007A
FILING DATE: March 8, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 5501969e
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-208-007A-8
Query Match 100.0%; Score 1806; DB 1; Length 331;
Best Local Similarity 100.0%; Pred. No. 4e-176;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MKRLVCVLLVCSAVALQHLKPTLDHWHWKKTYGQYKKEKNEARRLIWEKLIKQVM 60
|||||

Db 1 MKRLVCLLVCSAVALHDKDPTLDHMHLMKTYGQYKEKNEBAVRLIWEKNLKFVM 60
QY 61 LHNLESHMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMORNTTYSNPRILIPDSVD 120
Db 61 LHNLESHMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMORNTTYSNPRILIPDSVD 120
QY 121 WREKGCVEVKYQSGCGAFAVAGALEAOLKLTGKLVSLAQNLYDCSTERYGNKGC 180
Db 121 WREKGCVEVKYQSGCGAFAVAGALEAOLKLTGKLVSLAQNLYDCSTERYGNKGC 180
QY 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKRYATCSKYTELPYGRBDVLKE 240
Db 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKRYATCSKYTELPYGRBDVLKE 240
QY 241 AVANGPVSVDADARHPSPFLYRSQVYEPSCQVNVNNGVLVVGVDLNGKEVWLVKNSW 300
Db 241 AVANGPVSVDADARHPSPFLYRSQVYEPSCQVNVNNGVLVVGVDLNGKEVWLVKNSW 300
QY 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331
Db 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331

RESULT 2
US-09-701-685-2
; Sequence 2, Application US/09701685
; Patent No. 6387629
; GENERAL INFORMATION:
; APPLICANT: Schneider, Patrick
; APPLICANT: Yamamoto, Karen K.
; APPLICANT: French, Cynthia K.
; APPLICANT: Reprogen, Inc.
; TITLE OF INVENTION: Use of Cathepsin S in the Diagnosis and Treatment of
; FILE REFERENCE: 018002-001310US
; CURRENT FILING DATE: 2001-03-26
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: WO PCT/US99/12335
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: US 60/088,017
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-701-685-2

Query Match 100.0%; Score 1806; DB 2; Length 331;
Best Local Similarity 100.0%; Pred. No. 4e-176;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKRLVCLLVCSAVALHDKDPTLDHMHLMKTYGQYKEKNEBAVRLIWEKNLKFVM 60
Db 1 MKRLVCLLVCSAVALHDKDPTLDHMHLMKTYGQYKEKNEBAVRLIWEKNLKFVM 60
QY 61 LHNLESHMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMORNTTYSNPRILIPDSVD 120
Db 61 LHNLESHMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMORNTTYSNPRILIPDSVD 120
QY 121 WREKGCVEVKYQSGCGAFAVAGALEAOLKLTGKLVSLAQNLYDCSTERYGNKGC 180
Db 121 WREKGCVEVKYQSGCGAFAVAGALEAOLKLTGKLVSLAQNLYDCSTERYGNKGC 180
QY 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKRYATCSKYTELPYGRBDVLKE 240
Db 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKRYATCSKYTELPYGRBDVLKE 240
QY 241 AVANGPVSVDADARHPSPFLYRSQVYEPSCQVNVNNGVLVVGVDLNGKEVWLVKNSW 300
Db 241 AVANGPVSVDADARHPSPFLYRSQVYEPSCQVNVNNGVLVVGVDLNGKEVWLVKNSW 300

QY 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331
Db 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331

RESULT 3
US-09-290-586A-20
; Sequence 20, Application US/09290586A
; Patent No. 6800473
; GENERAL INFORMATION:
; APPLICANT: SANTAMARIA, Ignacio
; APPLICANT: VELASCO, Gloria
; APPLICANT: CAZORLA, Maite
; APPLICANT: FUEYO, Antonio
; APPLICANT: CAMPO, Elias
; APPLICANT: LOPEZ-OTIN, Carlos
; APPLICANT: AOKI, Takanori
; APPLICANT: IMATA, Kazuishi
; TITLE OF INVENTION: NOVEL HUMAN CATHEPSIN L2 PROTEIN, GENE ENCODING SAID
; FILE REFERENCE: 99-410A/MMC/01332
; CURRENT FILING DATE: US/09/290,586A
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: JP 10-172147
; PRIOR FILING DATE: 1998-06-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Human
US-09-290-586A-20

Query Match 100.0%; Score 1806; DB 2; Length 331;
Best Local Similarity 100.0%; Pred. No. 4e-176;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKRLVCLLVCSAVALHDKDPTLDHMHLMKTYGQYKEKNEBAVRLIWEKNLKFVM 60
Db 1 MKRLVCLLVCSAVALHDKDPTLDHMHLMKTYGQYKEKNEBAVRLIWEKNLKFVM 60
QY 61 LHNLESHMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMORNTTYSNPRILIPDSVD 120
Db 61 LHNLESHMGMSYDIGNMHLGDMTSEEVMSLSLRVPSOMORNTTYSNPRILIPDSVD 120
QY 121 WREKGCVEVKYQSGCGAFAVAGALEAOLKLTGKLVSLAQNLYDCSTERYGNKGC 180
Db 121 WREKGCVEVKYQSGCGAFAVAGALEAOLKLTGKLVSLAQNLYDCSTERYGNKGC 180
QY 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKRYATCSKYTELPYGRBDVLKE 240
Db 181 NGGFMTTAFQYIINDKIGDSASYPYKAMDLCQYDSKRYATCSKYTELPYGRBDVLKE 240
QY 241 AVANGPVSVDADARHPSPFLYRSQVYEPSCQVNVNNGVLVVGVDLNGKEVWLVKNSW 300
Db 241 AVANGPVSVDADARHPSPFLYRSQVYEPSCQVNVNNGVLVVGVDLNGKEVWLVKNSW 300
QY 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331
Db 301 GHNFGEGYIRMARKNHCGIASPSPYPEI 331

RESULT 4
US-08-860-255A-2
; Sequence 2, Application US/08860255A
; Patent No. 6274336
; GENERAL INFORMATION:
; APPLICANT: Abdel-Meguid, Sherin
; APPLICANT: Desjarlais, Renee
; APPLICANT: Janson, Cheryl
; APPLICANT: Smith, Ward
; APPLICANT: Zhao, Baoguang
; TITLE OF INVENTION: Method of Inhibiting Cathepsin K

FILE REFERENCE: P50574-X1
CURRENT FILING DATE: 1997-06-26
CURRENT FILING DATE: 1997-06-26
PRIOR APPLICATION NUMBER: 60/008,108
PRIOR FILING DATE: 1995-10-30
PRIOR APPLICATION NUMBER: 60/007,473
PRIOR FILING DATE: 1995-11-22
PRIOR APPLICATION NUMBER: 60/008,992
PRIOR FILING DATE: 1995-12-21
PRIOR APPLICATION NUMBER: 60/013,748
PRIOR FILING DATE: 1996-03-20
PRIOR APPLICATION NUMBER: 60/013,764
PRIOR FILING DATE: 1996-03-20
PRIOR APPLICATION NUMBER: 60/013,747
PRIOR FILING DATE: 1996-03-20
PRIOR APPLICATION NUMBER: 60/017,455
PRIOR FILING DATE: 1996-05-17
PRIOR APPLICATION NUMBER: 60/017,892
PRIOR FILING DATE: 1996-05-17
PRIOR APPLICATION NUMBER: 60/020,478
PRIOR FILING DATE: 1996-06-13
PRIOR APPLICATION NUMBER: 60/022,047
PRIOR FILING DATE: 1996-07-22
PRIOR APPLICATION NUMBER: 60/023,494
PRIOR FILING DATE: 1996-08-07
PRIOR APPLICATION NUMBER: 60/023,742
PRIOR FILING DATE: 1996-08-08
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 331
TYPE: PRT
ORGANISM: homo sapiens
US-08-860-255A-2

Query Match 99.2%; Score 1791; DB 2; Length 331;
Best Local Similarity 99.1%; Pred. No. 1.4e-174;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCLVLCSSAVALQHKDPTLDHMHLMKTYGKYKKEKNEBVARLLIWEKNLKFVM 60
DB 1 MKRLVCLVLCSSAVALQHKDPTLDHMHLMKTYGKYKKEKNEBVARLLIWEKNLKFVM 60
QY 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQORNTTYSNPNRILPDSVD 120
DB 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQORNTTYSNPNRILPDSVD 120
QY 121 WREKGVTEVYKQSCGACMAFSAVGALEAOLKLTGKLVLSAQNVLVDCSTERYGNKGC 180
DB 121 WREKGVTEVYKQSCGACMAFSAVGALEAOLKLTGKLVLSAQNVLVDCSTERYGNKGC 180
QY 181 NGGEMTTAFOYIINDKIGDSDASYPYKAMDLCQYDSKYRAATCSKYTELPYGREVDLKE 240
DB 181 NGGEMTTAFOYIINDKIGDSDASYPYKAMDLCQYDSKYRAATCSKYTELPYGREVDLKE 240
QY 241 AVANKGPVSVGVDAHPSEFLYRSGVYEPSCQONVNHGVLVVGDLNGKRYMLVKNSW 300
DB 241 AVANKGPVSVGVDAHPSEFLYRSGVYEPSCQONVNHGVLVVGDLNGKRYMLVKNSW 300
QY 301 GHNFGEGYIRMANRKNHCGIASPSPYPEI 331
DB 301 GHNFGEGYIRMANRKNHCGIASPSPYPEI 331

RESULT 5
US-08-915-095A-8
Sequence 8, Application US/08915095A
Patent No. 6383793
GENERAL INFORMATION:
APPLICANT: Haestings, et al.
TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: PFI07D4
CURRENT APPLICATION NUMBER: US/08/915, 095A

CURRENT FILING DATE: 1997-08-20
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 331
TYPE: PRT
ORGANISM: Homo sapiens
US-08-915-095A-8

Query Match 99.2%; Score 1791; DB 2; Length 331;
Best Local Similarity 99.1%; Pred. No. 1.4e-174;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCLVLCSSAVALQHKDPTLDHMHLMKTYGKYKKEKNEBVARLLIWEKNLKFVM 60
DB 1 MKRLVCLVLCSSAVALQHKDPTLDHMHLMKTYGKYKKEKNEBVARLLIWEKNLKFVM 60
QY 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQORNTTYSNPNRILPDSVD 120
DB 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQORNTTYSNPNRILPDSVD 120
QY 121 WREKGVTEVYKQSCGACMAFSAVGALEAOLKLTGKLVLSAQNVLVDCSTERYGNKGC 180
DB 121 WREKGVTEVYKQSCGACMAFSAVGALEAOLKLTGKLVLSAQNVLVDCSTERYGNKGC 180
QY 181 NGGEMTTAFOYIINDKIGDSDASYPYKAMDLCQYDSKYRAATCSKYTELPYGREVDLKE 240
DB 181 NGGEMTTAFOYIINDKIGDSDASYPYKAMDLCQYDSKYRAATCSKYTELPYGREVDLKE 240
QY 241 AVANKGPVSVGVDAHPSEFLYRSGVYEPSCQONVNHGVLVVGDLNGKRYMLVKNSW 300
DB 241 AVANKGPVSVGVDAHPSEFLYRSGVYEPSCQONVNHGVLVVGDLNGKRYMLVKNSW 300
QY 301 GHNFGEGYIRMANRKNHCGIASPSPYPEI 331
DB 301 GHNFGEGYIRMANRKNHCGIASPSPYPEI 331

RESULT 6
US-08-798-096-8
Sequence 8, Application US/08798096
Patent No. 6387682
GENERAL INFORMATION:
APPLICANT: Haestings, et al.
TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: PFI07D2
CURRENT APPLICATION NUMBER: US/08/798, 096
CURRENT FILING DATE: 1997-02-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 331
TYPE: PRT
ORGANISM: Homo sapiens
US-08-798-096-8

Query Match 99.2%; Score 1791; DB 2; Length 331;
Best Local Similarity 99.1%; Pred. No. 1.4e-174;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRLVCLVLCSSAVALQHKDPTLDHMHLMKTYGKYKKEKNEBVARLLIWEKNLKFVM 60
DB 1 MKRLVCLVLCSSAVALQHKDPTLDHMHLMKTYGKYKKEKNEBVARLLIWEKNLKFVM 60
QY 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQORNTTYSNPNRILPDSVD 120
DB 61 LHNLEHSMGMSYDLGNMHLGDMTSEEVMSLSLRVPSQORNTTYSNPNRILPDSVD 120
QY 121 WREKGVTEVYKQSCGACMAFSAVGALEAOLKLTGKLVLSAQNVLVDCSTERYGNKGC 180
DB 121 WREKGVTEVYKQSCGACMAFSAVGALEAOLKLTGKLVLSAQNVLVDCSTERYGNKGC 180
QY 181 NGGEMTTAFOYIINDKIGDSDASYPYKAMDLCQYDSKYRAATCSKYTELPYGREVDLKE 240

```
Db      181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDQCKQYDSKTRATCSKYTELPYGRBDVLKE 240
Oy      241 AVANKGPVSVGVDAHPSPFFLYRSGVYEPSCQOVNHNHGVLVVGYGDLNGKEXYMLVKNSW 300
Db      241 AVANKGPVSVGVDAHPSPFFLYRSGVYEPSCQOVNHNHGVLVVGYGDLNGKEXYMLVKNSW 300
Oy      301 GHNFGEEGYIRMARKNKGHCIGIASPPSYPEI 331
Db      301 GHNFGEEGYIRMARKNKGHCIGIASPPSYPEI 331
```

RESULT 7

```
US-08-798-095A-8
; Sequence 8, Application US/08798095A
; Patent No. 6423507
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D3
; CURRENT APPLICATION NUMBER: US/08/798,095A
; CURRENT FILING DATE: 1997-02-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-798-095A-8
```

```
Query Match      99.2%; Score 1791; DB 2; Length 331;
Best Local Similarity 99.1%; Pred. No. 1.4e-174;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      1 MKRLVCVLVCSAAYOLHKDPTLDHMHLMKKTGYKQYKENEAVRRLIWEKULKPFM 60
Db      1 MKRLVCVLVCSAAYOLHKDPTLDHMHLMKKTGYKQYKENEAVRRLIWEKULKPFM 60
Oy      61 LHNLEHSMGMSYDLGMNHLGDMTSEEVMSLSLRVPSQWORNITTYKSNPRLIPDSVD 120
Db      61 LHNLEHSMGMSYDLGMNHLGDMTSEEVMSLSLRVPSQWORNITTYKSNPRLIPDSVD 120
Oy      121 WREKCCVTEVYKQSCGACMAFSAVGALAEOLKLTGKLVSLISAQNLVDCSTEKYGNKC 180
Db      121 WREKCCVTEVYKQSCGACMAFSAVGALAEOLKLTGKLVSLISAQNLVDCSTEKYGNKC 180
Oy      181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
Db      181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
Oy      241 AVANKGPVSVGVDAHPSPFFLYRSGVYEPSCQOVNHNHGVLVVGYGDLNGKEXYMLVKNSW 300
Db      241 AVANKGPVSVGVDAHPSPFFLYRSGVYEPSCQOVNHNHGVLVVGYGDLNGKEXYMLVKNSW 300
Oy      301 GHNFGEEGYIRMARKNKGHCIGIASPPSYPEI 331
Db      301 GHNFGEEGYIRMARKNKGHCIGIASPPSYPEI 331
```

RESULT 8

```
US-09-953-956-8
; Sequence 8, Application US/09953956
; Patent No. 6475487
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D2D1
; CURRENT APPLICATION NUMBER: US/09/953,956
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: 09/219,441
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-953-956-8
```

```
Query Match      99.2%; Score 1791; DB 2; Length 331;
Best Local Similarity 99.1%; Pred. No. 1.4e-174;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      1 MKRLVCVLVCSAAYOLHKDPTLDHMHLMKKTGYKQYKENEAVRRLIWEKULKPFM 60
Db      1 MKRLVCVLVCSAAYOLHKDPTLDHMHLMKKTGYKQYKENEAVRRLIWEKULKPFM 60
Oy      61 LHNLEHSMGMSYDLGMNHLGDMTSEEVMSLSLRVPSQWORNITTYKSNPRLIPDSVD 120
Db      61 LHNLEHSMGMSYDLGMNHLGDMTSEEVMSLSLRVPSQWORNITTYKSNPRLIPDSVD 120
Oy      121 WREKCCVTEVYKQSCGACMAFSAVGALAEOLKLTGKLVSLISAQNLVDCSTEKYGNKC 180
Db      121 WREKCCVTEVYKQSCGACMAFSAVGALAEOLKLTGKLVSLISAQNLVDCSTEKYGNKC 180
Oy      181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
Db      181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
Oy      241 AVANKGPVSVGVDAHPSPFFLYRSGVYEPSCQOVNHNHGVLVVGYGDLNGKEXYMLVKNSW 300
Db      241 AVANKGPVSVGVDAHPSPFFLYRSGVYEPSCQOVNHNHGVLVVGYGDLNGKEXYMLVKNSW 300
Oy      301 GHNFGEEGYIRMARKNKGHCIGIASPPSYPEI 331
Db      301 GHNFGEEGYIRMARKNKGHCIGIASPPSYPEI 331
```

RESULT 9

```
US-08-553-125A-8
; Sequence 8, Application US/08553125A
; Patent No. 6475766
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D1
; CURRENT APPLICATION NUMBER: US/08/553,125A
; CURRENT FILING DATE: 1995-11-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-553-125A-8
```

```
Query Match      99.2%; Score 1791; DB 2; Length 331;
Best Local Similarity 99.1%; Pred. No. 1.4e-174;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      1 MKRLVCVLVCSAAYOLHKDPTLDHMHLMKKTGYKQYKENEAVRRLIWEKULKPFM 60
Db      1 MKRLVCVLVCSAAYOLHKDPTLDHMHLMKKTGYKQYKENEAVRRLIWEKULKPFM 60
Oy      61 LHNLEHSMGMSYDLGMNHLGDMTSEEVMSLSLRVPSQWORNITTYKSNPRLIPDSVD 120
Db      61 LHNLEHSMGMSYDLGMNHLGDMTSEEVMSLSLRVPSQWORNITTYKSNPRLIPDSVD 120
Oy      121 WREKCCVTEVYKQSCGACMAFSAVGALAEOLKLTGKLVSLISAQNLVDCSTEKYGNKC 180
Db      121 WREKCCVTEVYKQSCGACMAFSAVGALAEOLKLTGKLVSLISAQNLVDCSTEKYGNKC 180
Oy      181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
Db      181 NGGFMTTAFQYIINDKIGIDSDASYPKAMDLCQYDSKTRATCSKYTELPYGRBDVLKE 240
```

Qy	Db	Qy	Db
241	241	301	301
AVANKPEVSGVDARHPSEFLYRSGVYVEP	AVANKPEVSGVDARHPSEFLYRSGVYVEP	GHNFGEQYTRPARKNGHNGTASPSVPEI	GHNFGEQYTRPARKNGHNGTASPSVPEI
STQGVVNHGVLVWGQDNGKEVYLVN	STQGVVNHGVLVWGQDNGKEVYLVN	STQGVVNHGVLVWGQDNGKEVYLVN	STQGVVNHGVLVWGQDNGKEVYLVN
SNM	SNM	SNM	SNM
300	300	331	331
AVANKPEVSGVDARHPSEFLYRSGVYVEP	AVANKPEVSGVDARHPSEFLYRSGVYVEP	GHNFGEQYTRPARKNGHNGTASPSVPEI	GHNFGEQYTRPARKNGHNGTASPSVPEI
STQGVVNHGVLVWGQDNGKEVYLVN	STQGVVNHGVLVWGQDNGKEVYLVN	STQGVVNHGVLVWGQDNGKEVYLVN	STQGVVNHGVLVWGQDNGKEVYLVN
SNM	SNM	SNM	SNM
300	300	331	331

```

US-10-114-464-8
RESULT 10
US-10-114-464-8
Sequence 8, Application US/10114464
Patent No. 6680375
GENERAL INFORMATION:
APPLICANT: Haefliger, et al.
TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: PFI107D5
CURRENT APPLICATION NUMBER: US/10/114,464
CURRENT FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 08/553,125
PRIOR FILING DATE: 1995-11-07
PRIOR APPLICATION NUMBER: 08/208,007
PRIOR FILING DATE: 1994-03-08
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 331
TYPE: PRT
ORGANISM: Homo sapiens
US-10-114-464-8

```

Query Match	99.2%;	Score 1791;	DB 2;	Length 331;
Best Local Similarity	99.1%;	Pred. No. 1.4e-174;		
Matches 328;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0

Qy	1	MKLVLCVLLVCCSSAVQLHKDPTLPHNHMLAKKTYGKQYKPKNEAVARLLWEQNLKFFM	60
Db	1	MKLVLCVLLVCCSSAVQLHKDPTLPHNHMLAKKTYGKQYKPKNEAVARLLWEQNLKFFM	60
Qy	61	LHNLEISMGMHSDYLCGMNHLGDMTSEEWVLSLRSVPSQWRNIITYKSNPNRIILPDSVD	120
Db	61	LHNLEISMGMHSDYLCGMNHLGDMTSEEWVLSLRSVPSQWRNIITYKSNPNRIILPDSVD	120
Qy	121	WREKGCVTETKYGSGSCGACMAFSAVGALAEQKLKTGKLVLSLAQNLVDSCTEKYGNKGC	180
Db	121	WREKGCVTETKYGSGSCGACMAFSAVGALAEQKLKTGKLVLSLAQNLVDSCTEKYGNKGC	180
Qy	181	NGGPMTTAFQYIINDKIGDSDASYPYKAMDLCQYDSKYRAATCSKTELPYGRBDVLKE	240
Db	181	NGGPMTTAFQYIINDKIGDSDASYPYKAMDLCQYDSKYRAATCSKTELPYGRBDVLKE	240
Qy	241	AVANKRPVSGVDARHPSFFLYRSGYTYRPSCTQWVNHGVLVYSGDLNGEKETVLVKNW	300
Db	241	AVANKRPVSGVDARHPSFFLYRSGYTYRPSCTQWVNHGVLVYSGDLNGEKETVLVKNW	300
Qy	301	GHNFGEGYRMAARNKGHNCGIASPSPYPEI	331
Db	301	GHNFGEGYRMAARNKGHNCGIASPSPYPEI	331

RESULT 11
US-08-330-121B-4
Sequence 4, Application US/08330121B
Patent No. 5736357
GENERAL INFORMATION:
APPLICANT: Bromme, Dieter
APPLICANT: Okamoto, Kathleen
TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton
ADDRESSEE: & Hebert
STREET: Four Embarcadero Center, Suite 3400

```

1      CITY :   San Francisco
2      STATE:   California
3      COUNTRY: United States
4      ZIP:      94111-4187
5
6      COMPUTER READABLE FORM:
7
8      MEDIUM TYPE: Floppy disk
9
10     COMPUTER: IBM PC compatible
11     OPERATING SYSTEM: PC-DOS/MS-DOS
12     SOFTWARE: Patentin Release #1.0, Version #1.30
13
14     CURRENT APPLICATION DATA:
15
16     APPLICATION NUMBER: US/08/330,121B
17     FILING DATE: 27-OCT-1994
18     CLASSIFICATION: 435
19
20     ATTORNEY/AGENT INFORMATION:
21
22     NAME: Silva, Robin M.
23     REGISTRATION NUMBER: 38,304
24     REFERENCE/DOCKET NUMBER: A-60261/DJB/RMS
25     TELECOMMUNICATION INFORMATION:
26
27     TELEPHONE: (415) 781-1989
28
29     TELEFAX: (415) 398-3249
30
31     TELEX: 910 277299
32
33     INFORMATION FOR SEQ ID NO: 4:
34
35     SEQUENCE CHARACTERISTICS:
36
37     LENGTH: 331 amino acids
38     TYPE: amino acid
39
40     STRANDEDNESS: unknown
41
42     TOPOLOGY: unknown
43
44     MOLECULE TYPE: protein
45
46     US-08-330-121B-4

```

Query Match	98.8%;	Score 1784;	DB 1;	Length 331;
Best Local Similarity	99.1%;	Pred. No. 7.2e-174;		
Matches 328;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0

Qy	3	MKRLVCELLVSSSAVAQALHKDPTLDHNMILMKCTGQYKXKNEAARLLIWEZNLKLFVM	60
Db	1	MRLVCELLVSSSAVAQALHKDPTLDHNMILMKCTGQYKXKNEAARLLIWEZNLKLFVM	60
Qy	61	LHNLEHSMQMSHYDILGNNHLGDMTSEEWMSLMSLRVPSQORNIITYKSNPNRLLPDSVD	120
Db	61	LHNLEHSMQMSHYDILGNNHLGDMTSEEWMSLMSLRVPSQORNIITYKSNPNRLLPDSVD	120
Qy	121	WRKSGCVTEVTKQSGSGCAAFSAVGALEAOLKLTGTLVLSLAKQNLVDSCTERYGKGC	180
Db	121	WRKSGCVTEVTKQSGSGCAAFSAVGALEAOLKLTGTLVLSLAKQNLVDSCTERYGKGC	180
Qy	181	NGGFMTTAAQYIINDKGLDSDASYKKAMDLCQYDSTKTRAACTSKTTELPGHEVDLKE	240
Db	181	NGGFMTTAAQYIINDKGLDSDASYKKAMDLCQYDSTKTRAACTSKTTELPGHEVDLKE	240
Qy	241	AVANKPSPVGVDAHHPSEFLYRSVYVEPSCIQNVNHNGLVVGVDGLNGKEVYLVNNSV	300
Db	241	AVANKPSPVGVDAHHPSEFLYRSVYVEPSCIQNVNHNGLVVGVDGLNGKEVYLVNNSV	300
Qy	301	GNHFGEEGYIRMARKNKNGHGIASPPSYPEI	331
Db	301	GNHFGEEGYIRMARKNKNGHGIASPPSYPEI	331

RESULT 12
US-08-536-861-4
Sequence 4, Application US/08536861
Patent No. 6544767
GENERAL INFORMATION:
APPLICANT: Bromme, Dieter
APPLICANT: Okamoto, Kathleen
TITLE OF INVENTION: CATHPSIN O2 PROTEASE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobach, Test, Albritton & Hebert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California

COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/536,861
FILING DATE: 02-OCT-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-60261-1/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-536-861-4

Query Match 98.8%; Score 1784; DB 2; Length 331;
Best Local Similarity 99.1%; Pred. No. 7.2e-174;
Matches 328; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MKRLVCVLVCSAQAOLKDPPTLDHMHLMKTYGKQYKKEKNEBAVRLLIWEKNLKFVM 60
DB 1 MKRLVCVLVCSAQAOLKDPPTLDHMHLMKTYGKQYKKEKNEBAVRLLIWEKNLKFVM 60
QY 61 LHNLEHSGMHSYDIGNHLDGMTSEEVMSLSLRVPSQMRNITTYXNPNRIIPDSVD 120
DB 61 LHNLEHSGMHSYDIGNHLDGMTSEEVMSLSLRVPSQMRNITTYXNPNRIIPDSVD 120
QY 121 WREKCVTEVRYQSGCGACMAFSAVGALBAOLKLTGKLVLSAQNLYDCSTEKYGNKGC 180
DB 121 WREKCVTEVRYQSGCGACMAFSAVGALBAOLKLTGKLVLSAQNLYDCSTEKYGNKGC 180
QY 181 NGCFMTTAFQYIINDKGISDASYPYKAMDLCQYDSKTRATCSKYTELPYGRBDVLE 240
DB 181 NGCFMTTAFQYIINDKGISDASYPYKAMDLCQYDSKTRATCSKYTELPYGRBDVLE 240
QY 241 AVANKGPVSVGVDAHPSPFLYRSGVYVEPSCQYVNHGVLVVGVDLNGKEYWLVKNSW 300
DB 241 AVANKGPVSVGVDAHPSPFLYRSGVYVEPSCQYVNHGVLVVGVDLNGKEYWLVKNSW 300
QY 301 GHNFGEBGYIMARKNGHCGIASPSPYPEI 331
DB 301 GHNFGEBGYIMARKNGHCGIASPSPYPEI 331
```

RESULT 13

PCT-US95-13820-4

Sequence 4, Application PC/TUS9513820

GENERAL INFORMATION:

APPLICANT: Khepri Pharmaceuticals, Inc.

TITLE OF INVENTION: CATHEPSIN O2 PROTEASE

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13820
FILING DATE: 26-OCT-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US UNKNOWN
FILING DATE: 02-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/330,121
FILING DATE: 27-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: FP-60261-1-PC/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US95-13820-4

Query Match 98.8%; Score 1784; DB 4; Length 331;
Best Local Similarity 99.1%; Pred. No. 7.2e-174;
Matches 328; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MKRLVCVLVCSAQAOLKDPPTLDHMHLMKTYGKQYKKEKNEBAVRLLIWEKNLKFVM 60
DB 1 MKRLVCVLVCSAQAOLKDPPTLDHMHLMKTYGKQYKKEKNEBAVRLLIWEKNLKFVM 60
QY 61 LHNLEHSGMHSYDIGNHLDGMTSEEVMSLSLRVPSQMRNITTYXNPNRIIPDSVD 120
DB 61 LHNLEHSGMHSYDIGNHLDGMTSEEVMSLSLRVPSQMRNITTYXNPNRIIPDSVD 120
QY 121 WREKCVTEVRYQSGCGACMAFSAVGALBAOLKLTGKLVLSAQNLYDCSTEKYGNKGC 180
DB 121 WREKCVTEVRYQSGCGACMAFSAVGALBAOLKLTGKLVLSAQNLYDCSTEKYGNKGC 180
QY 181 NGCFMTTAFQYIINDKGISDASYPYKAMDLCQYDSKTRATCSKYTELPYGRBDVLE 240
DB 181 NGCFMTTAFQYIINDKGISDASYPYKAMDLCQYDSKTRATCSKYTELPYGRBDVLE 240
QY 241 AVANKGPVSVGVDAHPSPFLYRSGVYVEPSCQYVNHGVLVVGVDLNGKEYWLVKNSW 300
DB 241 AVANKGPVSVGVDAHPSPFLYRSGVYVEPSCQYVNHGVLVVGVDLNGKEYWLVKNSW 300
QY 301 GHNFGEBGYIMARKNGHCGIASPSPYPEI 331
DB 301 GHNFGEBGYIMARKNGHCGIASPSPYPEI 331
```

RESULT 14

US-10-010-577-2

Sequence 2, Application US/10010577

Patent No. 6784288

GENERAL INFORMATION:

APPLICANT: thurmond, robin l

APPLICANT: baker, sherry

TITLE OF INVENTION: polynucleotide and polypeptide sequences of monkey cathepsin s

FILE REFERENCE: ORT1457

CURRENT APPLICATION NUMBER: US/10/010,577

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patentin version 3.1

SEQ ID NO 2
LENGTH: 330
TYPE: PRT
ORGANISM: monkey
US-10-010-577-2

Query Match 94.2%; Score 1700.5; DB 2; Length 330;
Best Local Similarity 93.7%; Pred. No. 2.5e-165;
Matches 310; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

QY 1 MKRLVCVLVCSAAVQALHDKPTLDHMHMLKKTGYKQYKENEBAVRLLIWEKLLKFM 60
DB 1 MKQVLCVFCSSAAVQALHDKPTLDHMHMLKKTGYKQYKENEBAVRLLIWEKLLKFM 60
QY 61 LHNLEHSGMHSYDLGMNHLGDMTSEEVMSLSLRVPSQMRNITTKSNPRILLPDSVD 120
DB 61 LHNLEHSGMHSYDLGMNHLGDMTSEEVMSLSLRVPSQMRNITTKSNPRILLPDSVD 120
QY 121 WREKGCVEVYKQSGCGACMAFSAVGALBAQLKLTGKLVLSAQNLYDCSTKYGKGC 180
DB 121 WREKGCVEVYKQSGCGACMAFSAVGALBAQLKLTGKLVLSAQNLYDCS-EKYGNKGC 179
QY 181 NGGFMTTAFQYIIDNKGIDSDASYPKAMDLCQYDSKYRAATCSKYTELPYGRBDVKE 240
DB 180 NGGFMTTAFQYIIDNKGIDSEASYPKATDQCYDSKYRAATCSKYTELPYGRBDVKE 239
QY 241 AVANKGPVSGVDARHPSPFLYRSGVYEPSCQNVNNGVLVVGDLNGKEXYMLVKNSW 300
DB 240 AVANKGPVSGVDARHPSPFLYRSGVYEPSCQNVNNGVLVVGDLNGKEXYMLVKNSW 299
QY 301 GHNFGEGYIRMARNGNHCIGIASPSPYPEI 331
DB 300 GSNFGEGYIRMARNGNHCIGIASPSPYPEI 330

RESULT 15

US-10-010-580-2
Sequence 2, Application US/10010580
Patent No. 6780985
GENERAL INFORMATION:
APPLICANT: Chitmond, robin 1
APPLICANT: baker, sherry
TITLE OF INVENTION: polynucleotide and polypeptide sequences of canine cathepsin s
FILE REFERENCE: ORT-1458
CURRENT FILING DATE: 2001-11-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent version 3.1
SEQ ID NO 2
LENGTH: 331
TYPE: PRT
ORGANISM: canine
US-10-010-580-2

Query Match 88.5%; Score 1599; DB 2; Length 331;
Best Local Similarity 87.9%; Pred. No. 6e-155;
Matches 291; Conservative 19; Mismatches 21; Indels 0; Gaps 0;

QY 1 MKRLVCVLVCSAAVQALHDKPTLDHMHMLKKTGYKQYKENEBAVRLLIWEKLLKFM 60
DB 1 MKRLVCVLVCSAAVQALHDKPTLDHMHMLKKTGYKQYKENEBAVRLLIWEKLLKFM 60
QY 61 LHNLEHSGMHSYDLGMNHLGDMTSEEVMSLSLRVPSQMRNITTKSNPRILLPDSVD 120
DB 61 LHNLEHSGMHSYDLGMNHLGDMTSEEVMSLSLRVPSQMRNITTKSNPRILLPDSVD 120
QY 121 WREKGCVEVYKQSGCGACMAFSAVGALBAQLKLTGKLVLSAQNLYDCSTKYGKGC 180
DB 121 WREKGCVEVYKQSGCGACMAFSAVGALBAQLKLTGKLVLSAQNLYDCSTKYGKGC 180
QY 181 NGGFMTTAFQYIIDNKGIDSDASYPKAMDLCQYDSKYRAATCSKYTELPYGRBDVKE 240
DB 181 NGGFMTTAFQYIIDNKGIDSDASYPKAMDLCQYDSKYRAATCSKYTELPYGRBDVKE 240

DB 181 NGGFMTTAFQYIIDNKGIDSEASYPKAMNGRCRYDSKKRAATCSKYTELPYGRBDVKE 240
QY 241 AVANKGPVSGVDARHPSPFLYRSGVYEPSCQNVNNGVLVVGDLNGKEXYMLVKNSW 300
DB 241 AVANKGPVSGVDARHPSPFLYRSGVYEPSCQNVNNGVLVVGDLNGKEXYMLVKNSW 300
QY 301 GHNFGEGYIRMARNGNHCIGIASPSPYPEI 331
DB 301 GSNFGEGYIRMARNGNHCIGIASPSPYPEI 331

Search completed: January 10, 2006, 09:37:29
Job time: 46 secs

THIS PAGE BLANK (USPTO)